



[illegible]



BASE COUNT	1374 a	2545 c	2268 g	1140 t	5 others
ORIGIN	Chromosome q34.3.				
Query Match	9.0%; Score 27; DB 26; Length 7332;				
Best Local Similarity	73.7%; Pred. No. 4.19e-03;				
Matches	42; Conservative	0; Mismatches	15; Indels	0; Gaps	0;
Db	520	ACTGTCGGCAGAGTGTCAACGAGTGTGGCCAGAGCCAGGTTTGC	CGCCACGGA	576	
Qy	115	acctgcagtcagatgtgaatgagtgtgaataaaccccgccatcccaacacaga	171		
RESULT	7				
LOCUS	HS329A5	183338 bp	DNA	HTG	12-NOV-1998
DEFINITION	Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone 329A5; HTGS phase 1.				
ACCESSION	Z97832				
NID	g3355874				
KEYWORDS	HTG; HTGS_PHASE1.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
AUTHORS	1 (bases 1 to 183338)				
TITLE	Phillips.S.				
JOURNAL	Direct Submission				
COMMENT	Submitted (12-NOV-1998) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquires: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Jul 30, 1998 this sequence version replaced gi:3183868. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dJ329A5 Contig.ID: 01039 acc=Z97832 Length:				













\*\*\*\*\*  
 M P S R L A  
 \*\*\*\*\*  
 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.  
 Copyright (c) 1993-1998 University of Edinburgh, U.K.  
 Distribution rights by Oxford Molecular Ltd

Search\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
 Run on: Sat Apr 24 22:26:11 1999; MasPar time 60.26 Seconds  
 677.328 Million cell updates/sec  
 Tabular output not generated.  
 Title: >US-08-968-800-1  
 Description: (1-300) from US08968800.seq  
 Perfect Score: 299  
 N.A. Sequence: 1 ggctgagagaagaaacagcaa.....atagctgtgaagacacagaa 300  
 Comp: cagactctcttctgtgtt.....tategaacattctgtgtt

Scoring table: TABLE default  
 Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 188442 seqs, 58026449 bases x 2

Post-processing: Minimum Match 0%  
 Listing first 45 summaries

Database: n-geneseq32  
 1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
 8:part8 9:part9 10:part10 11:part11 12:part12 13:part13  
 14:part14 15:part15 16:part16 17:part17 18:part18  
 19:part19 20:part20 21:part21 22:part22 23:part23  
 24:part24 25:part25 26:part26 27:part27 28:part28  
 29:part29 30:part30 31:part31 32:part32 33:part33  
 34:part34 35:part35 36:part36 37:part37 38:part38  
 39:part39 40:part40

Statistics: Mean 7.783; Variance 4.414; scale 1.763

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	40	13.4	204 1 N81164	Base substituted E.co	1.16e-11
c 2	40	13.4	204 1 N81164	Base substituted E.co	1.16e-11
3	37	12.4	91 9 Q51746	Oligonucleotide probe	8.71e-10
c 4	37	12.4	91 9 Q51746	Oligonucleotide probe	8.71e-10
c 5	32	10.7	114 12 Q70469	Generic DNA sequence	9.60e-07
c 6	32	10.7	114 12 Q70469	Generic DNA sequence	9.60e-07
c 7	30	10.0	114 12 Q70465	Generic DNA sequence	1.46e-05
c 8	30	10.0	114 12 Q70467	Generic DNA sequence	1.46e-05
c 9	29	9.7	114 12 Q70470	Generic DNA sequence	5.60e-05
c 10	29	9.7	114 12 Q70466	Generic DNA sequence	5.60e-05
c 11	27	9.0	114 12 Q70465	Generic DNA sequence	7.85e-04
c 12	27	9.0	114 12 Q70468	Generic DNA sequence	7.85e-04
13	25	8.4	114 12 Q70469	Generic DNA sequence	1.03e-02

c 14	25	8.4	114 12 Q70467	Generic DNA sequence	1.03e-02
c 15	25	8.4	114 12 Q70472	Generic DNA sequence	1.03e-02
c 16	25	8.4	178 32 T76405	Human endothelin-1 an	1.03e-02
c 17	25	8.4	720 1 Q05947	Sequence encoding thr	1.03e-02
c 18	25	8.4	726 1 Q05948	Sequence encoding thr	1.03e-02
c 19	25	8.4	747 1 Q05945	Sequence encoding thr	1.03e-02
c 20	25	8.4	759 1 Q05945	Sequence encoding thr	1.03e-02
c 21	25	8.4	825 21 T14656	cdNA encoding human r	1.03e-02
c 22	25	8.4	825 1 Q05971	Sequence encoding thr	1.03e-02
c 23	25	8.4	831 1 Q05973	Sequence encoding thr	1.03e-02
c 24	25	8.4	1425 3 Q20811	Encodes truncated hum	1.03e-02
c 25	25	8.4	1482 16 Q59556	Encodes thrombomodul	1.03e-02
c 26	25	8.4	1491 21 T14654	cdNA encoding a human	1.03e-02
c 27	25	8.4	1529 16 Q59558	Encodes thrombomodul	1.03e-02
c 28	25	8.4	1545 2 Q10435	Soluble thrombomodul	1.03e-02
c 29	25	8.4	1656 2 Q13406	Thrombin-binding subs	1.03e-02
c 30	25	8.4	1680 16 Q59557	Encodes thrombomodul	1.03e-02
c 31	25	8.4	1680 4 Q52579	Recombinant thrombin	1.03e-02
c 32	25	8.4	1946 21 T12792	Human thrombomodulin	1.03e-02
c 33	25	8.4	2011 6 Q36458	Human thrombomodulin	1.03e-02
c 34	25	8.4	2313 1 Q05944	Sequence encoding thr	1.03e-02
c 35	25	8.4	2463 3 Q20810	EcoRI fragment encodi	1.03e-02
c 36	25	8.4	3373 2 N80892	Sequence encoding new	1.03e-02
c 37	25	8.4	3573 1 N81065	Sequence of human thr	1.03e-02
c 38	25	8.4	3640 1 N82026	Cloned p2.1 insert en	1.03e-02
c 39	25	8.4	6977 8 Q48213	Hind III - Not I thro	1.03e-02
c 40	24	8.0	114 12 Q70466	Generic DNA sequence	3.63e-02
c 41	24	8.0	114 12 Q70470	Generic DNA sequence	3.63e-02
c 42	24	8.0	114 12 Q70473	Generic DNA sequence	3.63e-02
c 43	24	8.0	114 12 Q70471	Generic DNA sequence	3.63e-02
c 44	23	7.7	81 21 T13611	DC43 TSAR library gen	1.25e-01
c 45	23	7.7	140 32 T76368	Human IL-8 receptor-a	1.25e-01

ALIGNMENTS

RESULT 1  
 ID N81164 standard; DNA; 204 BP.  
 AC N81164;  
 DT 08-NOV-1990 (first entry)  
 DE Base substituted E.coli beta-galactosidase alpha-fragment.  
 KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.  
 OS Escherichia coli.  
 FH key Location/Qualifiers  
 FT misc\_feature 19..69  
 FT primer\_bind 187..204 /function=multiple cloning site  
 FT /\*tag= b  
 PN EP-285123-A.  
 PD 05-MAY-1988.  
 PF 30-MAR-1988; 105163.  
 PR 03-APR-1987; US-034819.  
 PA (SUSO) SUOMEN SOKRI OY.  
 PI Lehtovaara P, Koovles J, Koivula A, Bamford J, Reinikainen T;  
 DR WPI: 88-279927/40.  
 PT Introducing random point mutations into nucleic acids -  
 PT by prepn of single stranded template, annealing a primer, elongation,  
 PT misincorporation, completion of molecules and screening.  
 PS Disclosure: p; English.  
 CC Random point mutations were introduced into the alpha fragment of  
 CC single beta-galactosidase. The wild type sequence was obtained as a  
 CC single stranded template and an oligonucleotide was hybridised to  
 CC it to generate a popn of DNA molecules which terminate at all  
 CC possible nucleotide positions within a specified region. The  
 CC variable 3' ends generated in this way are used as primers for  
 CC reverse transcriptase. Nucleotides are misincorporated by the  
 CC transcriptase and the molecules are completed to forms that can be  
 CC amplified and then expressed in a suitable host-vector system.  
 CC The sequence covers all 176 diff base substitutions, most of which  
 CC occurred singularly in any given mutant.  
 CC See also P80375.  
 SQ sequence 204 BP; 21 A; 47 C; 17 G; 11 T; 108 Others;

```

PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 12.4%; Score 37; DB 9; Length 91;
Best Local Similarity 2.0%; Pred. No. 8.71e-10;
Matches 1; Conservative 43; Mismatches 7; Indels 0; Gaps 0;

Db 10 gssvhsyyvvhvshhsvhhvhhvhsvvvhhvhhvhhvhhvhyvsv 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 61 ggtgagtcgtggcgaccacaatgcagatgctttccaggataccagg 111

RESULT 4
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DE 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
KW ss.
OS Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
DR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
PS Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 12.4%; Score 37; DB 9; Length 91;
Best Local Similarity 2.0%; Pred. No. 8.71e-10;
Matches 1; Conservative 43; Mismatches 7; Indels 0; Gaps 0;

Db 10 gssvhsyyvvhvshhsvhhvhhvhsvvvhhvhhvhhvhhvhyvsv 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Cp 98 ggaagcatctgcattgtttggtccacgcactcaccacattacatcca 48

RESULT 5
ID Q70469 standard; DNA; 114 BP.
AC Q70469;
DR 07-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
DE TSAR; totally synthetic affinity reagent; synthetic; binding domain;
DE effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH key
FT misc_feature 55..60
Location/Qualifiers

```



PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PS comprising a binding domain and an effector domain  
 PS Disclosure; Page 35; 255pp; English.  
 CC Q70465 is a generic DNA sequence used to generate random TSAR (Totally  
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
 CC represented as follows: X(NNB)6(TGC)(NNB)12(NNB)14(TGC)(NNB)3Y. X  
 CC and Y are flanking restriction sites (X is not the same as Y) that are  
 CC not specified further. Other generic sequences are shown in Q70466-68.  
 CC Other specific peptides generated by these generic sequences are shown in  
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,  
 CC comprising at least two functional regions - a binding domain with  
 CC affinity for a ligand and a second effector peptide portion that is  
 CC chemically or biologically active. They may further comprise a linker  
 CC peptide between the 2 domains. The oligonucleotides are also designed so  
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned  
 CC in, or flanking, the unpredicted or variant residues. These residues  
 CC confer some degree of conformational rigidity to the peptides. The TSARs  
 CC or compsns. comprising a TSAR binding domain can be used in vivo to  
 CC deliver a chemically or biologically active moiety, eg. metal ion,  
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the  
 CC cell. They can also replace the function of macromolecules, eg.  
 CC monoclonal or polyclonal antibodies and therefore circumvent the need  
 CC for complex methods of hybridoma formation or in vivo antibody  
 CC production. The TSARs are easily characterised and have designed  
 CC activity allowing direct and rapid detection in a screening process.  
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 10.0%; Score 30; DB 12; Length 114;  
 Best Local Similarity 3.6%; Pred. No. 1.46e-05;  
 Matches 4; Conservative 31; Mismatches 77; Indels 0; Gaps 0;  
 Db 3 bnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62  
 Cp 188 ccgtgtgtattcacacatctgtgtgcatgcccggggtttcattccacattccaca 129  
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114  
 Cp 128 tcttgactgcagggtttcccggtgtatctcgtggaagcattcgtcattgttg 77

RESULT 8  
 ID Q70467 standard; DNA; 114 BP.  
 AC Q70467;  
 DT 05-APR-1995 (first entry)  
 DE Generic DNA sequence to generate a random TSAR peptide library.  
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
 KW effector domain; concatenated heterofunctional protein; linker;  
 KW direct; rapid; detection; screening; treatment; generic; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 55..60  
 FT /\*tag= a  
 FT /note= "this sequence represents 'Z'; Z can be a  
 FT sequence of 6, 9 or 12 nucleotides (see  
 FT comments)."  
 PN WO9418318-A.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U00977.  
 PR 01-FEB-1994; US-013416.  
 PR 30-DEC-1993; US-176500.  
 PR 31-JAN-1994; US-189331.  
 PA (UINC-) UNIV NORTH CAROLINA.  
 PI Fowlkes DM, Kay BK;  
 DR WPI; 94-279739/34.  
 DR P-PSDB; R65153.

PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PS comprising a binding domain and an effector domain  
 PS Disclosure; Page 35; 255pp; English.  
 CC Q70467 is a generic DNA sequence used to generate random TSAR (Totally  
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
 CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)1Y. X

CC and Y are flanking restriction sites (X is not the same as Y) that are  
 CC not specified further. Other generic sequences are shown in Q70466-68.  
 CC Other specific peptides generated by these generic sequences are shown in  
 CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,  
 CC comprising at least two functional regions - a binding domain with  
 CC affinity for a ligand and a second effector peptide portion that is  
 CC chemically or biologically active. They may further comprise a linker  
 CC peptide between the 2 domains. The oligonucleotides are also designed so  
 CC that the expressed peptide contains 2 or 4 cysteine residues positioned  
 CC in, or flanking, the unpredicted or variant residues. These residues  
 CC confer some degree of conformational rigidity to the peptides. The TSARs  
 CC or compsns. comprising a TSAR binding domain can be used in vivo to  
 CC deliver a chemically or biologically active moiety, eg. metal ion,  
 CC radioisotope, peptide, toxin or enzyme, to the specific target or on the  
 CC cell. They can also replace the function of macromolecules, eg.  
 CC monoclonal or polyclonal antibodies and therefore circumvent the need for  
 CC complex methods of hybridoma formation or in vivo antibody production.  
 CC The TSARs are easily characterised and have designed activity allowing  
 CC direct and rapid detection in a screening process.  
 SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;  
 Query Match 10.0%; Score 30; DB 12; Length 114;  
 Best Local Similarity 3.6%; Pred. No. 1.46e-05;  
 Matches 4; Conservative 31; Mismatches 77; Indels 0; Gaps 0;  
 Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62  
 Cp 194 tagcttcggtgtattcacacatctgtgtgcatgcccggggtttcattccacattca 135  
 Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114  
 Cp 134 ttacatttgcagggtttcccggtgtatctcgtggaagcattcgtcatt 83

RESULT 9  
 ID Q70470 standard; DNA; 114 BP.  
 AC Q70470;  
 DT 10-APR-1995 (first entry)  
 DE Generic DNA sequence to generate a random TSAR peptide library.  
 KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;  
 KW effector domain; concatenated heterofunctional protein; linker;  
 KW direct; rapid; detection; screening; treatment; generic; ss.  
 OS Synthetic.  
 FH Key Location/Qualifiers  
 FT misc\_feature 55..60  
 FT /\*tag= a  
 FT /note= "encoded by Z (see comments)"  
 PN WO9418318-A.  
 PD 18-AUG-1994.  
 PF 01-FEB-1994; U00977.  
 PR 01-FEB-1994; US-013416.  
 PR 30-DEC-1993; US-176500.  
 PR 31-JAN-1994; US-189331.  
 PA (UINC-) UNIV NORTH CAROLINA.  
 PI Fowlkes DM, Kay BK;  
 DR WPI; 94-279739/34.  
 DR P-PSDB; R58378.

PT Identifying proteins or peptide(s) which bind a ligand - by  
 PT screening a recombinant vector library expressing fusion proteins  
 PS comprising a binding domain and an effector domain  
 PS Disclosure; Page 36; 255pp; English.  
 CC Q70470 is a generic DNA sequence used to generate random TSAR (Totally  
 CC Synthetic Affinity Reagents) peptides. This generic formula can also be  
 CC represented as follows: X(NNB)4(CAC)(NNB)4(CAC)(NNB)8Z(NNB)6(CAC)(NNB)8  
 CC -(CAC)2(NNB)Y. X and Y are flanking restriction sites (X is not the same  
 CC as Y) that are not specified further. The peptides generated by this and  
 CC other generic sequences (Q70471-73) have invariant histidine residues  
 CC incorporated into variant sequences. TSARs are concatenated  
 CC heterofunctional proteins or peptides, comprising at least two functional  
 CC regions - a binding domain with affinity for a ligand and a second  
 CC effector peptide portion that is chemically or biologically active. They  
 CC may further comprise a linker peptide between the 2 domains. The TSARs  
 CC or compsns. comprising a TSAR binding domain can be used in vivo to



QY 174 tgtgaatacacacggaagctacaagtgccttctcctcagtgggccacatgctcatgcaga 233  
Db 63 bnbnnbn 114  
QY 234 tgctacgtgtgtaactcnaggacatgtgccatgataaactgcagtatagc 285

```

RESULT 12
ID Q70468 standard; DNA; 114 BP.
AC Q70468;
DE DT 05-APR-1995 (first entry)
DE DE Generic DNA sequence to generate a random TSAR peptide library.
DE KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
DE KW effector domain; concatenated heterofunctional protein; linker;
DE KW direct; rapid; detection; screening; treatment; generic; ss.
DE KW synthetic.
DE OS
DE FT
DE FT Key Location/Qualifiers
DE misc_feature 55..60 /tag= a
DE note= "this sequence represents 'Z'; Z can be a
sequence of 6, 9 or 12 nucleotides (see
comments."

```

Comments)

W09419318-A. 18-AUG-1994. 000977. U009977. 01-FEB-1994; 01-FEB-1993; US-013416. 30-DEC-1993; US-176500. 31-JAN-1994; US-189331. (UINC-) UNIV NORTH CAROLINA. Fowlkes DM, Kay BK; P-PSDB; R65154. Identifying proteins or peptide(s) which bind a ligand - by screening a recombinant vector library expressing fusion proteins Disclosure; Page 35; 255pp; English. Q70468 is a generic DNA sequence used to generate random TSAR (Totally Synthetic Affinity Reagents) peptides. This generic formula can also be represented as follows: X(NNB)11(TGC)(NNB)62(TGC)(NNB)10Y. X and Y are flanking restriction sites (X is not the same as Y) that are other specified further. Other generic sequences are shown in Q70466-68. R65151-54. TSARS are concatenated heterofunctional proteins or peptides, comprising at least two functional regions - a binding domain with affinity for a ligand and a second effector peptide portion that is chemically or biologically active. They may further comprise a linker peptide between the 2 domains. The oligonucleotides are also designed so that the expressed peptide contains 2 or 4 cysteine residues positioned in, or flanking, the unpredictable or variant residues. These residues confer some degree of conformational rigidity to the peptides. The TSARS or compns. comprising a TSAR binding domain can be used in vivo to deliver a chemically or biologically active moiety, eg. metal ion, radioisotope, peptide, toxin or enzyme, to the specific target or on the cell. They can also replace the function of macromolecules, eg. monoclonal or polyclonal antibodies and therefore circumvent the need for complex methods of hybridoma formation or in vivo antibody production. The TSARS are easily characterised and have designed activity allowing direct and rapid detection in a screening process. Sequence 114 Bp: 9 A. 2 G.

```

Query Match          9.0%; Score 27; DB 12; Length 114;
Quest Local Similarity 3.6%; Pred. No. 7.85e-04;
Matches             4; Conservative 29; Mismatches 79; Indels 0; Gaps 0;

3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnnn 62
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
174 tgtgaatacacaggaagctacaagtgcttttgcctcagtggccacatgctcatgcaga 233
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 tgtcactgtgtgaactcnagacatgtgcctatgataaactgcagctatgc 285
  : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 13
ID Q70469 standard; DNA; 114 BP.
AC Q70469;
DE 07-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
DE TSAR; totally synthetic affinity reagent, synthetic binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
KW Synthetic.
OS Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents '2'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
W094183118-A.
18-AUG-1994.
PDN 18-AUG-1994.
PPF 01-FEB-1994; U000977.
PPF 01-FEB-1994; US-013416.
PPR 30-DEC-1993; US-176500.
PPR 31-JAN-1994; US-189331.
PPR (UYNC-) UNIV NORTH CAROLINA.
PPR Fowlkes DM, Kay BK;
PI WPI: 94-279739/34.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PT Disclosure; Page 35; 255pp; English.
TS

```

Q70469 is a generic DNA sequence used to generate random TSAR peptide  
 This generic formula can be represented as follows: X(TGC)(NNB)10-  
 (TGC)(NNB)6Z(NNB)2(TGC)(NNB)14(TCY). X and Y are flanking restriction  
 sites (X is not the same as Y) that are not specified further. This  
 sequence generates peptides that are cloverleaf in structure. Other  
 generic sequences are shown in Q70465-68. Other specific peptides  
 generated by these generic sequences are shown in R65150-54. TSARs are  
 concatenated heterofunctional proteins or peptides, comprising at least  
 two functional regions - a binding domain with affinity for a ligand and  
 a second effector peptide portion that is chemically or biologically  
 active. They may further comprise a linker peptide between the 2 domains.  
 The oligonucleotides are also designed so that the expressed peptide  
 contains 2 or 4 cysteine residues positioned in, or flanking, the  
 unpredicted or variant residues. These residues confer some degree of  
 conformational rigidity to the peptides. The TSARs or compns. comprising  
 a TSAR binding domain can be used in vivo to deliver a chemically or  
 biologically active moiety, eg. metal ion, radioisotope, peptide, toxin  
 or enzyme, to the specific target or on the cell. They can also replace  
 the function of macromolecules, eg. monoclonal or polyclonal antibodies  
 and therefore circumvent the need for complex methods of hybridoma  
 formation or in vivo antibody production. The TSARs are easily  
 characterised and have designed activity allowing direct and rapid  
 detection in a screening process.  
 Sequence 114 bp.

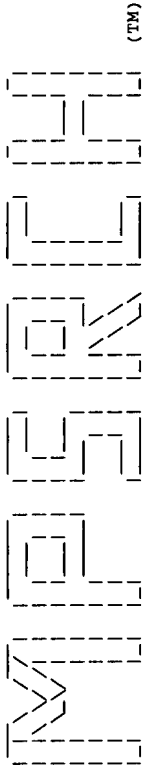
	Query Match	8.4%	Score 25;	DB 12:	Length 114;
	Best Local Similarity	6.1%	Pred. No. 1.03E-02;		
	Matches	7;	Conservative	27; Mismatches	80; Indels
					0; Gaps
Db	1	tgcnnbnbnbnbnbnbnbnbnbnbnbnbtgcnbnbnbnbnbnbnbnbnbnbnbnbn	60		
QY	172	tgtgtgaatacacagcgaaagtctacagtgtctttgctcagtgggcacatgctcatgcc	231		
Db	61	nubnbnbtgcnbn	114		
QY	232	gatgctacgttgtgtgaactcnaggacatgtggccatgataaacgttcagtcatagc	285		

RESULT	14
ID	Q70467 standard; DNA; 114 BP.
AC	Q70467;
DE	05-APR-1995 (first entry)
DEFN	Generic DNA sequence to generate a random
DESC	TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KEYW	TSAR; totally synthetic affinity reagent; synthetic; binding domain;





\*\*\*\*\*



Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

psrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Sat Apr 24 22:17:30 1999; MasPar time 484.95 Seconds  
Tabular output not generated. 1107.807 Million cell updates/sec

Title: >US-08-968-800-1  
Description: (1-300) from US08968800.seq  
Perfect Score: 299  
N.A. Sequence: 1 gctgtgagaagaacacgcaa.....atagctgtgaacacagaa 300  
Comp: cgcacacctttctgttgcgtt.....tatcgacacttctgtt

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-est56  
Database: genbank-est109  
1:em\_est1 2:em\_gss1 3:em\_gss2 4:em\_gss3  
5:gb\_est1 6:gb\_est10 7:gb\_est11 8:gb\_est12 9:gb\_est13  
10:gb\_est14 11:gb\_est15 12:gb\_est16 13:gb\_est17  
14:gb\_est18 15:gb\_est19 16:gb\_est2 17:gb\_est20  
18:gb\_est21 19:gb\_est3 20:gb\_est4 21:gb\_est5 22:gb\_est6  
23:gb\_est7 24:gb\_est8 25:gb\_est9 26:gb\_gss1 27:gb\_gss2  
28:gb\_gss3 29:gb\_gss4

Statistics: Mean 9.757; Variance 1.776; scale 5.495

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Description	Pred. No.
1	113	37.8	420 18 AA913032	0134a06.s1 Soares_NFL_ 1.05e-185
2	54	18.1	252 12 AA754459	97SN1787 Rice Immature 9.83e-62
3	52	17.4	252 12 AA754459	97SN1787 Rice Immature 8.07e-58
4	43	14.4	247 12 AA754458	97SN1784 Rice Immature 1.16e-40
5	41	13.7	247 12 AA754458	97SN1784 Rice Immature 5.67e-37
6	39	13.0	334 10 AA614649	np48a02.s1 NCI-CGAP.Br 2.46e-33
7	27	9.0	2275 11 AF034173	Homo sapiens ntcon2 co 4.91e-13
8	26	8.7	311 9 W09270	ma05h07.r1 Soares mous 1.68e-11
9	26	8.7	365 21 W82677	me98h07.r1 Soares mous 1.68e-11
10	26	8.7	473 21 AA028399	mi21e01.r1 Soares mous 1.68e-11
11	26	8.7	483 7 AA497481	vh29c03.r1 Soares mous 1.68e-11
12	26	8.7	554 12 AA793395	vn52d12.r1 Barstead mo 1.68e-11
13	24	8.0	2275 11 AF034173	Homo sapiens ntcon2 co 1.54e-08

RESULT	1	AA913032	420 bp	mrna	EST	24-SEP-1998
LOCUS	1	ol34a06.s1 Soares_NFL.T.GBC.S1	Homo sapiens	CDNA clone		
DEFINITION		IMAGE:1525330 3' similar to SW:FBNI_MOUSE Q61554	FIBRILLIN 1			
ACCESSION		AA913032				
NID		93052424				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1 (bases 1 to 420)				
AUTHORS		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP).				
COMMENT		Tumor Gene Index Unpublished (1997)				
FEATURES		Location/Qualifiers				
source		1. .420				
		/organism="Homo sapiens"				
		/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19w, testis NHT, and B-cell NCI-CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made				

ALIGNMENTS

RESULT	1	AA913032	420 bp	mrna	EST	24-SEP-1998
LOCUS	1	ol34a06.s1 Soares_NFL.T.GBC.S1	Homo sapiens	CDNA clone		
DEFINITION		IMAGE:1525330 3' similar to SW:FBNI_MOUSE Q61554	FIBRILLIN 1			
ACCESSION		AA913032				
NID		93052424				
KEYWORDS		EST.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		1 (bases 1 to 420)				
AUTHORS		Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
TITLE		NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.				
JOURNAL		National Cancer Institute, Cancer Genome Anatomy Project (CGAP).				
COMMENT		Tumor Gene Index Unpublished (1997)				
FEATURES		Location/Qualifiers				
source		1. .420				
		/organism="Homo sapiens"				
		/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19w, testis NHT, and B-cell NCI-CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made				

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1169 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 68.  
Location/Qualifiers  
1. .420  
/organism="Homo sapiens"  
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19w, testis NHT, and B-cell NCI-CGAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaudo.  
 /db\_xref="taxon:9606"  
 /clone\_image="1525330"  
 /clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
 /lab\_host="DH10B"  
 /lab\_host="DH10B"

BASE COUNT 96 a 102 c 124 g 98 t  
 ORIGIN

Query Match 37.8%; Score 113; DB 18; Length 420;  
 Best Local Similarity 70.7%; Pred. No. 1.05e-185;  
 Matches 191; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Db 110 GGCACAGTGTGACCCGTGTGCAACACGATGCAACATGTTGAATGTATCGGCAAA 169  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Qy 21 gggagtctgtgaagctacatcgaaacctggtgaattgtgagtcgctgggaccaa 80  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 170 CAAGTGCAGTGTATCTCTGTTATGCTGGAAACCTGTATCAAGATCTAAATGAGTG 229  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 81 caaatgcagatgcttccaggatcacaccgggaaacctgcagtcgaagtgaatgagtg 140  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 230 TGGCTGAAGCCCGCCCTGTAAACACAGGTCATGACACTTACCGGACGCTACAAGTG 289  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Qy 141 tggatgaacccgcacatgccacacagatgtgtgaatacacacggaagctacaagt 200  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 290 CTACTGTCTCAACGATATATGCTATCGGATGTTCTCTCAAGTGCCCTGACCTG 349  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Qy 201 ctttgcctcagtgccacatgctcagtcagtcagtcagtcagtcagtcagtcagtc 260  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Db 350 CTCCATGGCAACATGCTAGTATGCTGTGA 379  
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
 Qy 261 tgcatgataaactgtcagtagctgtga 290

RESULT 2  
 LOCUS AA754459 252 bp mRNA  
 DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa  
 CDNA clone 97SN1787, mRNA sequence.  
 ACCESSION AA754459  
 NID 92801165  
 KEYWORDS EST.  
 SOURCE rice.  
 ORGANISM Oryza sativa  
 Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;  
 Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;  
 Poales; Poaceae; Oryza.  
 REFERENCE 1 (bases 1 to 252)  
 AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.  
 TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Eun M.Y.  
 Department of Cytoogenetics  
 National Inst. of Agri. Sci. and Tech, RDA  
 Suwon, Kyunggido, Korea  
 Tel: 82 331 290 0301  
 Fax: 82 331 290 0307  
 Email: myeun@sun20.asti.re.kr  
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr  
 Seq primer: M13 Reverse Primer.  
 Location/Qualifiers  
 1..252  
 /organism="Oryza sativa"  
 /cultivar="Milyang23"  
 /note="Vector: pBluescript SK(+); Site1: EcoRI; Site2: XhoI; Directional CDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."  
 /db\_xref="taxon:4530"

FEATURES  
 source  
 Location/Qualifiers  
 1..252  
 /organism="Oryza sativa"  
 /cultivar="Milyang23"  
 /note="Vector: pBluescript SK(+); Site1: EcoRI; Site2: XhoI; Directional CDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."  
 /db\_xref="taxon:4530"

/clone="97SN1787"  
 /clone\_lib="Rice Immature Seed Lambda ZAPII CDNA Library"  
 /tissue\_type="Immature Seed"  
 /dev\_stage="5 days after pollination"  
 /lab\_host="E. coli SOLR"  
 BASE COUNT 5 a 21 c 12 g 35 t 179 others  
 ORIGIN

Query Match 18.1%; Score 54; DB 12; Length 252;  
 Best Local Similarity 13.0%; Pred. No. 9.83e-62;  
 Matches 33; Conservative 121; Mismatches 95; Indels 4; Gaps 4;

Db 1 HWDCMTMTVTRGCCCCBAMTTSYBCHGNVYCVASHGNTYMHVHCTBRGTHCDCKNVNW 60  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Cp 267 catggcacatgctcngaggttcacacacgtagcatctgcatgcatgcatgcatgcatg 209  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Db 61 STMTGTVNVBNYSGDWHYWBVENTKVDGNHTRCSRWBVRMAHYHDYTNCBBYNNND 120  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Cp 208 ggcacacgcatctgtagcttcctggtattcacacatctggttgcatggcgggggtt 149  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Db 121 YHWHBMBYBBTCTMCMWBHNTKCTASGWHSTNYDVKSSNTWGTBSYDKSMH 180  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Cp 148 tcattccacact-cat-tcacatctgactgcaggtttcccggtgtatctctggaagca 91  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Db 181 GTWCBBVYHTKVTTRATRSYTCVRKYCVMMTKVVKVHVBB-GCHBTDCKKTM 239  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Cp 90 tctgcattgttggccccacgcatccacacacattacatccaggttcgcatgtagcttc 31  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Db 240 WMTNKHVMTSTTD 252  
 :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: |||| :::: ||||  
 Cp 30 acagactcccttg 18

RESULT 3  
 LOCUS AA754459 252 bp mRNA  
 DEFINITION 97SN1787 Rice Immature Seed Lambda ZAPII CDNA Library Oryza sativa  
 CDNA clone 97SN1787, mRNA sequence.  
 ACCESSION AA754459  
 NID 92801165  
 KEYWORDS EST.  
 SOURCE rice.  
 ORGANISM Oryza sativa  
 Eukaryotae; Viridiplantae; Charophyta/Embryophyta group;  
 Embryophyta; Tracheophyta; seed plants; Magnoliophyta; Liliopsida;  
 Poales; Poaceae; Oryza.  
 REFERENCE 1 (bases 1 to 252)  
 AUTHORS Nahm,B.H., Kim,J.K., Cheong,J.J., Kim,S.I., Hahn,T.R., Moon,E.P., Kim,W.T., Kim,W.Y., Yang,M.S., Park,R.D., Sohn,U.I., Kang,K.Y., Lee,M.C. and Eun,M.Y.  
 TITLE Large-scale Sequencing Analysis of ESTs from Rice Immature Seed  
 JOURNAL Unpublished (1998)  
 COMMENT Contact: Eun M.Y.  
 Department of Cytoogenetics  
 National Inst. of Agri. Sci. and Tech, RDA  
 Suwon, Kyunggido, Korea  
 Tel: 82 331 290 0301  
 Fax: 82 331 290 0307  
 Email: myeun@sun20.asti.re.kr  
 Submitted by Baek Hie Nahm, Dept of Biological Science, Myongji University, Yongin, Korea. 449-728 bhnam@bioserver.myongji.ac.kr  
 Seq primer: M13 Reverse Primer.  
 Location/Qualifiers  
 1..252  
 /organism="Oryza sativa"  
 /cultivar="Milyang23"  
 /note="Vector: pBluescript SK(+); Site1: EcoRI; Site2: XhoI; Directional CDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."  
 /db\_xref="taxon:4530"  
 /clone="97SN1787"  
 /clone\_lib="Rice Immature Seed Lambda ZAPII CDNA Library"  
 /tissue\_type="Immature Seed"

FEATURES  
 source  
 Location/Qualifiers  
 1..252  
 /organism="Oryza sativa"  
 /cultivar="Milyang23"  
 /note="Vector: pBluescript SK(+); Site1: EcoRI; Site2: XhoI; Directional CDNA library inserted into lambda ZAPII vector at 5' end with EcoRI and 3' end with Xho I site."  
 /db\_xref="taxon:4530"  
 /clone="97SN1787"  
 /clone\_lib="Rice Immature Seed Lambda ZAPII CDNA Library"  
 /tissue\_type="Immature Seed"



QY 83 aatgcagatgcttccagatatacaccggaacacccctgcagcgaagtgtgaatgagtgtg 142  
 Db 173 GGGCNRKVTYGSBRYRCGVNVMVVRTSMWTDKSTKMSMDMSRRSRVHYGRWMBNKK 230  
 QY 143 gaatgaaccccgccgcatgccaacacagatgtgtgaatacacacggaagctacaagt 200

RESULT 6  
 LOCUS AA614649 334 bp mRNA EST 16-OCT-1997  
 DEFINITION np48a02.s1 NCI-CGAP\_Br1.1 Homo sapiens cDNA clone IMAGE:1129514,  
 mRNA sequence.  
 ACCESSION AA614649  
 NID 92466845  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 334)  
 AUTHORS NCI-CGAP  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT  
 Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 Cloning Distribution: Washington University Genome Sequencing Center  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1982 Std Error: 0.00  
 Seq primer: -40m3 fwd. Et from Amersham

High quality sequence stop: 318.  
 Location/Qualifiers

## FEATURES

source

1..334  
 /organism="Homo sapiens"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker; 1st strand cDNA was prepared from pooled bulk  
 breast tumor tissue, and was then primed with a Not I -  
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pT73  
 vector. Library is not normalized. (The normalized  
 version of this library is NCI-CGAP\_Br2.) Library was  
 constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:9606"  
 /clone.lib="NCI-CGAP\_Br1.1"  
 /sex="female, pooled"  
 /tissue\_type="breast"  
 /lab\_host="DH10B"  
 BASE COUNT 98 a 60 c 65 g 111 t  
 ORIGIN

Query Match 13.0%; Score 39; DB 10; Length 334;  
 Best Local Similarity 68.9%; Pred. No. 2.46e-33;  
 Matches 71; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Db 144 TTCTACCTTGATACAGGTTTTCACGATACACGATGACATGCTGCTTTGGC 203  
 Cp 134 ttcaacttgatgcaggttttcccggtgtatctggaagcatctgctattgttgg 75

Db 204 CGATACATTCACCATGTTTGCATCGGTGGCACACAGCTT 245  
 Cp 74 cccgcgactcaccacactatcatcaggttcgcattgactt 32

RESULT 7  
 LOCUS AF034173 2275 bp mRNA EST 22-DEC-1997  
 DEFINITION Homo sapiens ntcon2 contig mRNA, partial sequence.  
 ACCESSION AF034173  
 NID 92707735  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2275)  
 AUTHORS Tripodis, N. and Ragoussis, J.  
 TITLE Generation of a transcription map in the region immediately  
 centromeric to human MHC across the 6p21.2-6p21.3 chromosomal  
 boundary  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 2275)  
 AUTHORS Tripodis, N. and Ragoussis, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-NOV-1997) Division of Medical and Molecular Genetics,  
 Guy's Hospital, 7th floor, Guy's Tower, London SE1 9RT, UK  
 FEATURES  
 Location/Qualifiers  
 1..2275  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /map="6p21.3"  
 /clone="ntcon2 contig"  
 /tissue\_type="liver; brain"  
 /dev\_stage="fetus"  
 /note="similar to Br140"  
 BASE COUNT 438 a 619 c 470 g 599 t 149 others  
 ORIGIN

Query Match 9.0%; Score 27; DB 11; Length 2275;  
 Best Local Similarity 10.6%; Pred. No. 4.91e-13;  
 Matches 10; Conservative 50; Mismatches 34; Indels 0; Gaps 0;

Db 1476 AAMWYKRWKRGRKRMTGMVKRYRAMMAMCMMACWYKMRGKKWKYKRYK 1535  
 QY 164 aacacagatgctggaatacacagcgaagtacagtgcttgcagtgccacatgc 223

Db 1536 KYTSTYKGSRSRYWTTTYYWYWCWCTSMKKSAS 1569  
 QY 224 tcacgcagatgctacgtgtgtgaactcnaggac 257

RESULT 8  
 LOCUS W09270 311 bp mRNA EST 01-OCT-1997  
 DEFINITION ma05h07.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 303709 5',  
 similar to SW.TLD.DROME P25723 DORSAL-VENTRAL PATTERNING TOLLOID  
 PROTEIN PRECURSOR ;, mRNA sequence.  
 ACCESSION W09270  
 NID 91283586  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 311)  
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,  
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,  
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,  
 Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and  
 Waterston, R.  
 TITLE The WashU-HMI Mouse EST Project  
 JOURNAL Unpublished (1996)  
 COMMENT  
 Contact: Marra M/Mouse EST Project  
 WashU-HMI Mouse EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108



```

/organism="Mus musculus"
/strain="C3H"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; lst strand cDNA
was primed with a Not I -- oligo(dT) primer [5',
TGTTCAGCAATCGATGGAGCGCGCCCTTTTTTTTTTTTTTTTTT
3']; Double-stranded cDNA was ligated to Eco RI adaptors
[AAATTCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead. The C2C12 cell line
(available from ATCC, catalog # CRL-1772) differentiates

```



Search completed: Sat Apr 24 22:25:49 1999  
Job time : 499 secs.



\*\*\*\*\*  
MIPESH (TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

msrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
on: Thu Apr 29 03:05:32 1999; MasPar time 2494.84 Seconds  
1525.983 Million cell updates/sec  
Tabular output not generated.

Title: >US-08-968-800-2  
Description: (1-1611) from US08968800.seq  
Perfect Score: 1607  
N.A. Sequence: 1 ggcgtgagaagaacagca.....aattactagctgaaaaattg 1611  
Comp: ccgacctttcttctgctgt.....ttaagatcgacttttaac

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 602357 seqs, 1181590623 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: emb157  
1:em\_ba 2:em\_fun 3:em\_htg 4:em\_hum1 5:em\_hum2 6:em\_in  
7:em\_om 8:em\_ov 9:em\_ov 10:em\_pat 11:em\_ph 12:em\_pl  
13:em\_ro 14:em\_un 15:em\_v1

Database: genbank110  
16:gb\_ba1 17:gb\_ba2 18:gb\_htg 19:gb\_in 20:gb\_om 21:gb\_ov  
22:gb\_pat 23:gb\_ph 24:gb\_pl 25:gb\_pi2 26:gb\_pr1  
27:gb\_pr2 28:gb\_pr3 29:gb\_ro 30:gb\_st 31:gb\_sts 32:gb\_sy  
33:gb\_un 34:gb\_v1

Statistics: Mean 11.423; Variance 5.514; scale 2.072

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	108	6.7	7218	22	Sequence 14 from patent	1.81e-60
2	41	2.6	215	22	Sequence 5 from patent	1.24e-09
3	37	2.3	2070	19	Caenorhabditis elegans	4.99e-07
4	37	2.3	2139	19	Caenorhabditis elegans	4.99e-07
5	37	2.3	10772	19	Drosophila melanogaste	4.99e-07
6	35	2.2	7218	22	Sequence 14 from patent	9.13e-06
7	36	2.2	9662	26	Homo sapiens fibrillin	2.15e-06
8	36	2.2	9940	26	HsFIBRLLN	2.15e-06
9	35	2.2	10772	19	Drosophila melanogaste	9.13e-06
10	36	2.2	74371	27	Homo sapiens chromosom	2.15e-06
11	33	2.1	2156	26	HsFIBUC	1.56e-04
12	33	2.1	2349	26	HsFIBUC	1.56e-04
13	33	2.1	2359	26	Human fibrillin-ID mRNA,	1.56e-04

14	33	2.1	2525	26	HSFIBUB	H.sapiens mRNA for fib	1.56e-04
15	32	2.0	9830	29	MMU22493	Mus musculus fibrillin	6.25e-04
16	32	2.0	11724	29	AF007248	Mouse fibrillin mutant f1	6.25e-04
17	32	2.0	11953	29	MUSFBN1A	Mouse fibrillin (Fbn-1	6.25e-04
18	30	1.9	215	22	I28278	Sequence 5 from patent	9.42e-03
19	30	1.9	965	22	AR024229	Sequence 22 from patent	9.42e-03
20	31	1.9	1561	26	HUMCLNXX	Human calnexin mRNA, c	2.45e-03
21	31	1.9	1288	26	HUMCALNEXI	Human calnexin mRNA, c	2.45e-03
22	31	1.9	3881	26	HUMCALNEXI	Human calnexin mRNA, c	2.45e-03
23	31	1.9	4117	26	HUMIP90	Homo sapiens integral	2.45e-03
24	30	1.9	74371	27	AC005369	Homo sapiens chromosome	9.42e-03
25	29	1.8	151	22	MMUIDEX12	M. musculus nid gene (e	3.53e-02
26	29	1.8	985	22	AR024229	Sequence 22 from patent	3.53e-02
27	29	1.8	2338	21	AF051400	Gallus gallus fibrulin-	3.53e-02
28	29	1.8	2709	29	MMMB90B	M. musculus (isolate MK	3.53e-02
29	29	1.8	3346	21	AF051399	Gallus gallus fibrulin-	3.53e-02
30	29	1.8	4171	29	MMNIDOG	Mouse mRNA for nidogen	3.53e-02
31	29	1.8	4633	21	AF027596	Danio rerio tolloid mr	3.53e-02
32	29	1.8	5959	29	MMENACT	Mouse mRNA for entacti	3.53e-02
33	29	1.8	9859	29	MUSFBN2	Mus musculus fibrillin	3.53e-02
34	29	1.8	10172	26	HSU03272	Human fibrillin-2 mRNA	3.53e-02
35	29	1.8	216021	27	HUAC004787	Homo sapiens Chromosom	3.53e-02
36	29	1.8	238893	18	HS90L6	Human DNA sequence ***	3.53e-02
37	28	1.7	1548	24	GMNGL16	Soybean Ngm-16 gene co	1.29e-01
38	28	1.7	2823	26	HSFIB5	H.sapiens mRNA for fib	1.29e-01
39	28	1.7	8950	20	BOVXAAA	Bos taurus mRNA, compl	1.29e-01
40	27	1.7	35593	19	CELC56E5	Caenorhabditis elegans	4.60e-01
41	27	1.7	37001	27	AC005513	Homo sapiens chromosome	4.60e-01
42	27	1.7	125536	18	AC003021	*** SEQUENCING IN PROG	4.60e-01
43	28	1.7	158427	27	AC002553	Homo sapiens chromosome	1.29e-01
44	27	1.7	183338	18	HS329A5	Human DNA sequence ***	4.60e-01
45	27	1.7	216021	27	HUAC004787	Homo sapiens Chromosom	4.60e-01

ALIGNMENTS

RESULT 1	I66494	7218 bp	DNA	PAT	23-DEC-1997
LOCUS	Sequence 14 from patent US 5670367.				
DEFINITION	166494				
ACCESSION	166494				
NID	9274471				
KEYWORDS	Unknown.				
SOURCE	Unknown.				
ORGANISM	Unclassified.				
REFERENCE	1 (bases 1 to 7218)				
AUTHORS	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.				
TITLE	Recombinant fowlpox virus				
JOURNAL	Patent: US 5670367-A 14 23-SEP-1997;				
FEATURES	Location/Qualifiers				
source	1..7218				
BASE COUNT	1944 a 1491 c 1486 g 1929 t				
ORIGIN	/organism="unknown"				
Query Match 6.7%; Score 108; DB 22; Length 7218;					
Best Local Similarity 1.6%; Pred. No. 1.81e-60;					
Matches 6; Conservative 231; Mismatches 129; Indels 0; Gaps 0;					
Db	1075	Y	Y	Y	Y
Cp	1031	t	t	t	t
Db	1135	Y	Y	Y	Y
Cp	971	c	c	c	c
Db	1195	Y	Y	Y	Y
Cp	911	t	t	t	t
Db	1255	Y	Y	Y	Y

Cp 851 ttcccttttttacctccatgagagttccgcctctggaacatctctcttcattagtggaag 792  
 Db 1315 YY 1374  
 Cp 791 ggcgcagctccacttaggtaggagctgctggtctctgggtaacatttttaatt 732  
 Db 1375 YY 1434  
 Cp 731 ttgctctcttttctgctgtttttgtgagcaagcaactctgtattctgtcttgatg 672  
 Db 1435 GTACCA 1440  
 Cp 671 gtacca 666

RESULT 2 I28278 215 bp DNA PAT 30-OCT-1996  
 LOCUS Sequence 5 from patent US 5569830.  
 DEFINITION I28278  
 ACCESSION g1819054

WORDS  
 SOURCE Unknown.  
 ORGANISM Unknown.

REFERENCE 1 (bases 1 to 215)

AUTHORS Bennett, A., Labavitch, J.M., Powell, A. and Stotz, H.

TITLE Plant inhibitors of fungal polygalacturonases and their use to

JOURNAL control fungal disease

FEATURES Patent: US 5569830-A 5 29-OCT-1996;

Location/Qualifiers

1. .215

/organism="unknown"

BASE COUNT 15 a 8 c 25 g 26 t 141 others

ORIGIN

Query Match 2.6%; Score 41; DB 22; Length 215;  
 Best Local Similarity 16.3%; Pred. No. 1.24e-09;  
 Matches 34; Conservative 82; Mismatches 91; Indels 2; Gaps 2;

Db 7 SSSSVVSRASCNDKAKKGGNTTSWTTDCNRTWGVCDTDTTVRVNNDGHNKYSANY 66

Cp 343 ccaggcgagctcctgagtaggacacagcagctggtgcccctctctctgcttcacagc 284

Db 67 NYGGNNVGAAKTHYYHTNVSGADSKVTVDSYNASGTSSSNGTGNRSGADSYGSSKTA 126

Cp 283 tatactgacagtttatcatggca-catgctcctngagttcacacagctagcatctggcag 225

Db 127 MTSNRTGTANNVDSRNNGDASVGSNDKNTKKHAKNSADGKVGSKNNGDRNNRYGTGK 186

Cp 224 agcagtggccactgagggcaaaagcactgttagctccgtgtgtattc-acacatctgtg 166

Db 187 SNVSNCGGNKRDVSSYANNKCGSSCT 215

Cp 165 ttggcatgcccggggtttcattccacact 137

RESULT 3 AF051401 2070 bp mRNA INV 02-AUG-1998  
 LOCUS Caenorhabditis elegans fibulin-1 isoform D precursor (FBLN1) mRNA,  
 DEFINITION alternatively spliced product, complete cds.  
 ACCESSION AF051401  
 NID g3372520

WORDS

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE 1 (bases 1 to 2070)

AUTHORS Barth, J.L., Argraves, K.M., Roark, E.F., Little, C.D. and

TITLE Isolation of chicken and nematode fibulin-1 homologs and

JOURNAL characterization of the nematode fibulin-1 gene

REFERENCE 2 (bases 1 to 2139)

Unpublished

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## FEATURES

## source

## 1. .2070

## /organism="Caenorhabditis elegans"

## /db\_xref="taxon:6239"

## /chromosome="IV"

## /map="between elt-1 and daf-10"

## /note="partial sequence obtained from EST clone yk11b1"

## 1. .2070

## /gene="FBLN1"

## 1. .2070

## /gene="FBLN1"

## /note="extracellular matrix protein"

## /codon\_start=1

## /product="fibulin-1 isoform D precursor"

## /db\_xref="PID:g3372521"

## /translation="MRICLLFLAFLVAETFAELTRCAGGTRHFKNSNTGSSIKSEK

## TSMTCORAAISCLLRLDNACDSTGDIKKEEPCSNINILGGKKECCDCCLLAK

## DILNNPECPVAPVGFSAQLRSFNKCCNGDIEITHASEIITGRPLNDPVLHGDRC

## SSGHEHLCHDRGKVEGSCRSRGLADPGMACVDIDECATLMDCCLESQRLNTPGS

## FKCIRTLCGTYAMDSERCDVDECNLGHSDGGLYQCRNTOGSRCDKAKKCEDV

## ELQNPMTGECTSITCPNGYYPKNGMNDIDECVTGHNCGAGGECVNTPGSPRCQKGN

## LCANGYEVNGATGCEDEVNECOQGVGSGMECINLPGYKCKGPGYFNDKAKKEDV

## DECINKFAGHVCDSLAECLINTGSEFCCKKPGFOLASDGRCEDEYNETGTGTAACEQK

## VNIPIGYQICDRGALFGPDGKEDIDECISWAGSNDRLCMGCIINTKGSYLQCCPC

## GYKIQPDGRTCDVDECAEGCAGSKDVCVNTLGSFKCHSDCFTNTHSLNKNQIA

## DGYSICKVCTEDTECLNHTREVLVQFRAVPSLKTILISPIEVSRIVTHMGVPSVDY

## NLDTVGORHFRIVQERNIGIVQLVKPISGPTVETIKVNIHTKSTGVLAFNEALIEI

## SVSKYPT"

## BASE COUNT 640 a 337 c 493 g 600 t

## ORIGIN

## Query Match

## Best Local Similarity 2.3%; Score 37; DB 19; Length 2070;

## Matches 79; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

## Db 544 TCTGGATTGCACTTGGCACCAGATGGTGTGTGTAGATATTGATGTGTCACACA 603

## Qy 328 tcaggactccgctggcccaaatggaagactgtctagatattgtagatgtgcctct 387

## Db 604 CTGATGACCAATGTTTGGAACTCTCAACGATGTTTGAATCTCTCTGGAAGCTTTAAATGC 663

## Qy 388 ggtaaagtcactgtccctacaatcgaagtgtggaacattggaagactactactgc 447

## Db 664 A 664

## Qy 448 a 448

## RESULT 4

## LOCUS

## DEFINITION

## ACCESSION

## NID

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE



Matches	20;	Conservative	63;	Mismatches	46;	Indels	0;	Gaps	0;
---------	-----	--------------	-----	------------	-----	--------	----	------	----

[illegible]

<b>RESULT</b>	6		
<b>LOCUS</b>	I66494	7218 bp	DNA
<b>DEFINITION</b>	Sequence 14 from patent US 5670367.		PAT
<b>ACCESSION</b>	Z66494		23-DEC-1997

92/24471

SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source

.  
Unknown.  
Unknown.  
Unclassified.  
1 (bases 1 to 7218)  
Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.  
Recombinant fowlpox virus  
Patent: US 5670367-A 14 23-SEP-1997;  
Location/Qualifiers  
1. 7218

BASE COUNT	/organism="unknown"		
ORIGIN	1944 a	1491 c	1929 t
	1486 g.		368 others

Query Match 2.2%; Score 35; DB 22; Length 7218;  
Best Local Similarity 2.1%; Pred. No. 9.13e-06;  
Matches 3; Conservative 83; Mismatches 54; Indels 0; Gaps 0

[illegible]

	7	HUMFIBRLN	9662 bp	mRNA	PRI
RESULT		DEFINITION		fibrillin mRNA,	complete cds.
FOCUS		L13923			
ACCESSION		J308745			
ID		Fibrillin.			
EYWORDS		human.			
SOURCE		Homo sapiens			
ORGANISM					

REFERENCE	Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS	1 (bases 1 to 9662) Pereira,L.V., D'Alessio,M., Ramirez,F., Lynch,J., Sykes,B., Pangilinan,T. and Bonadio,J. Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome Hum. Mol. Genet. 2, 961-968 (1993) 93372860
TITLE	
JOURNAL	
MEDLINE	
FEATURES	Location/Qualifiers 1..9662 /organism="Homo sapiens" /db_xref="taxon:9606" 1..133
source	
5'UTR	

5'UTR

**CDS**

134. .8749

```

/standard_name="Marfan syndrome gene"
/codon_start=1
/product="fibrillin"
/db_xref="PID:g306746"
/translation="MRRGRLLLEALGFTVLLASYSHGADANLEAGNVETRASRAKR
RGGGHDALKGNVGVSRVNAVCCPGWITLPGNQCIPICRHSOGDGFCSRPNMCTC
PSGOIAPSGSRSIOHCNIRCMNMGSSDHDLCQKQYIGTHCGVPSGSGLNGGRC
VAPNRCACTYGTGQCERDRTGCFVTIVSNMQCGQLSIGVCPKQLCATVGRWAG
HPCEMCAQHPHRCRGTIPNRTAGACVDQCAIPLGCGGNCINVSYSFECKCPAG
HKLNEVSCQEDIDECSTIPGTCEGECTNTVSSYFKCPGCFYVSPGTGTRDVRPG
YCYTALTNGRCSNOLPQSIKTMQCCDAGRQSPGVTAPEMCPTRATEDFNKLCSVP
MVIPEPVPPEPLPGIPVLPVPGPFGPQIAPVRPEVEILYPSRPPRVLPNVNT
GYCOLVRYLCOGRCIPIPGSYRCECNAGFQLDREGCIDVDECKNFCAAGCETNNQ
GSYTCQCRAGYQSIULTRECHDIDECQLQNGRCINNGRCINTDGSFYHCNAGFHVTRD
GKNCEDMDECSIRNCLNMGMINEDGSKICKPQOLASDGRYCKDINCECTPGICH
NRGNVNTDGSYRBCFPGLAVGLDGRVCVDTHMSTRCTYGVYKRGQCIKPLGAVTKSE
CCCATYFAGCEPCQCPAONSAEVOALCSSPGMWTSAQNSIDECALDIPCPNGICE
NLRTGYCICNGSYEVDSYTKNCVDINECVLNSLLDNGQCRNTPGSFVCTCPKGFYI
KPDLTCTEDIDECSSPCINGCAAGSPGSFICESSSETLPTTKTICETIKGTCTWT
VIDGRCEININGATLKSQCSSLGAAGSPCTLCQVDPICGKYSIRKGTQCEDIDEC
EYRPGVNGCLVNAWGTFCQCPQSGMTLDATGRICLDIRLETGCFRLYDEECTPLIA
GRHMDACCCVGAAGTATECECECPRENTPEYELCPRGPFATKEITNGEPPKFDIN
EKEMPISLCTHGRKNTIGSKFKCDGSFALDSEERNCTDIDECRIPLDGRGCGOVN
TPQDFECKDEGVESGPMMKMNCMDIDECQDPLLCRPGVCHTGETSYRCECPGPHQL
SPNISACTIDNECELSAHLCPNGRCVNLIKYQACANPGVHSTPDRFLCYVDIDECIS
NGGCEPACTINSESYECSCQPGFALMPDQSCSTDIDECENPNIDCGGQCTNIPEYR
CLCIDYFMASEDNMTKTVVNECDLNPICISGTCENTKGSFICHDMMYSGKKKCTGC
DNTNEEIGAHNCKYAGVNTAGTSFKSCSPGWIGDIKCTDLDECSNPTHMCQSQA
DKNTINGSLCKEYGTGDTCTDLDECSNLLNCGQCLDNLAPGGYRCECDMGFV
PSTADGACEDIDECSLPNICVCTCHNLPOLFRCEBGLYELDRSGNCTDVENCLDP
TTCISGNVNTPGSYICDPPPELAPVTRGVCTRSGNCLDIRPRDNGDTACSNE
IGVGSKACCCSLKANGTPELCPNCPAVNTSEYKILCPGGEFRPNITVILEDIDEC
QELPLCGQKINTFGSCQCRPTGLYINLEDTRVCDVNECETPGICPGTCVNTYG
NATYICDPDYMVGNGNCDMRSLCYRNYADNCTDCELLFNMTMKCCSXYNG
RAWNRCEQCPISPDEFAITLGSQRPGFVIDYITGLPVIDIDECREIPYGVENGVCIN
MWSFRCECPGVFFYHDKLLVEDIDECQNGVPCORNAECINTAGSTRCDCKPGYRFT
STGQCDNRNECEIPINICSHGQCIDVGSFYCLCHTGFKTNDDOTMCLDINECEDAC
NGTCTNTIGSYNCRNCHGFIISHNNDKIDVDECSNGLNCRNGQINTVGSFQQC
NRYEYVADPGRCTDINECLNECPALCPAGTCQNLDSYRICKPGYSLOAKCEDIDE
CVREEPICALGTCSTNTEGFKLCEPTEFSLSRSGRCQDLRMSYKAKREBGGKSSPK
SRNKHCEALCKEGMGDPCELPTGEDFAFRQICPYGSGIIVGDDPAMVMDCKP
EDPVCKHGQICINTDGSYRCECPFGYTLAGNECVTDDECYVNGPCNGTKCNVIGGFEC
TCEEGFEPGPMTCEDINECAPNLLCAFCVNTYGSYCKCPGVTLREDRMDCKDE
DBCEBGNKCTEKOMEKNLIGTYMCIGPGYQRRPDGECVDENECCTKPGTCENGR
CLNTGYSYTCENCDGTFASPNODECLDNREGYCTFVLQNMCOIGSSNPNVPTKSECC
CDGGRWGPHCEBICPFQGTVAFKLCPHGRGEMTFAGADIECKYIHDCVRNGBCVNDR
GSYHCKITGTYPDITGYSVDLNECNQAPKPFCTKNTEGYSQCSQPKGYILQEDG
RSKDLDECATKHQNCQFLCVNTIGGFTKCPPGTOHTHSCIDNNECTSDINLCSGK
LQIOTOWNOCVDNECLASHGGASCHNTLGSYKCMCPAGFYQFYESGQCDINEGCS
AQAPCSYCSNTEGYLCGCPPGFRIQGGHCVSGMGMRGPNPEVPVSGEMDNLSLP
EACVECKINGTPKRRKRRNSETDASNIEDQSETEANYSLASVDVEKTAIFAENLISH
VSNVRIELLEPLATLTHNRNLIETESNEDGFFKINQEGISLYHFTKKKPVAGTYS
LQISSTPLYKKELNQLBKDYLDKVLSELGDLNKKMKIQLLH"
9662
polya_site
2619 a 2161 c 2490 g 2392 t
BASE COUNT
ORIGIN

```

**Query Match**            2.28;       Score 36;     DB 26;     Length 9662;  
**Best Local Similarity** 66.48;  
**Matches**              73;     Conservation  
                                 0 %;

[illegible]

## RESULT 8





```

/rpt_family="Alu"
27191..27477
/rpt_family="Alu"
27774..28057
/rpt_family="Alu"
28040..28066
/note="A127"
/rpt_type-tandem
/rpt_unit=A
28712..28930
/rpt_family="MER20"
complement(28769..28838)
/note="GRAIL 2 excellent exon, frame 0"
repeat_region
28987..29214
/rpt_family="Alu"
join(29485..29559,29598..29779)
/note="95% identity dbSTS:G14522 (SHGC-11312)"
misc_feature
29495..29976
/standard_name="AI027942"
/note="100% identity EST ov84al0.x1"
complement(30401..30536)
/note="GRAIL 2 excellent exon, frame 2"
repeat_region
complement(30682..30733)
/rpt_family="MIR"
complement(31573..31724)
/note="GRAIL 2 excellent exon, frame 1"
misc_feature
complement(32159..32232)
/note="GRAIL 2 excellent exon, frame 2"
repeat_region
32388..32488
/rpt_family="MLT1"
complement(33670..33785)
/rpt_family="Alu"
complement(34021..34144)
/note="GRAIL 2 excellent exon, frame 2"
misc_feature
complement(35238..35331)
/note="GRAIL 2 excellent exon, frame 1"
repeat_region
36392..36663
/rpt_family="Alu"
complement(36901..37222)
/note="100% identity EST ou55c09.x1"
misc_feature
36901..37164
/db_xref="dbEST:AI025011"
STS
/standard_name="STSC-9983"
/db_xref="dbSTS:G26554"
complement(37595..37654)
/note="GRAIL 2 excellent exon, frame 0"
misc_feature
complement(join(38063..38218,38462..38578,38741..38995,39071..39205,39532..39630,39935..40048,40300..40410,40503..40661,41868..41972,42103..42225,42492..42569,44379..44507))
/standard_name="histidyl-tRNA synthetase"
/note="78%-100% protein identity GenPept:U18937"
misc_feature
complement(38069..38215)
/note="GRAIL 2 excellent exon, frame 0"
misc_feature
complement(38462..38578)
...
Note: remainder of annotations omitted.

Query Match      2.2%; Score 36; DB 27; Length 74371;
Best Local Similarity 15.8%; Pred.No. 2.15e-06;
Matches 16; Conservative 55; Mismatches 29; Indels 1; Gaps 1;

Db 15931 SRSKSRGWRGYSRWKKYRCAMWMTCKSKKWCWSYRMKCYSCSYCCSSGKKYWCRC 15990
      : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Cp 913 cattcttcagggtcttctct-tctctttctcatctcaagccctcttctctct 855

Db 15991 MYWYTCYSYKYWMSYCYCTCTSGWRWMSKGRSWMYASR 16031
      : : : : : : : : : : : | : : : : : | : : : : : |
Cp 854 tcattccctttttacctcatgacagattctccgcctctgga 814

```

..  
Note: remainder of annotations omitted.

```

RESULT 11
LOCUS HSF1BUC 2156 bp RNA PRI 21-AUG-1995
DEFINITION H.sapiens mRNA for fibulin-1 C.
ACCESSION X53743
NID g31418
KEYWORDS fibulin-1 C; glycoprotein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2156)
AUTHORS Argraves, W.S.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1990) Argraves W.S., American Red Cross, 15601
Crabbs Branch Way, Rockville, MD 20855, USA
REFERENCE 2 (bases 1 to 2156)
AUTHORS Argraves, W.S., Tran, H., Burgess, W.H. and Dickerson, K.
TITLE Fibulin is an extracellular matrix and plasma glycoprotein with
repeated domain structure
JOURNAL J. Cell Biol. 111 (6 Pt 2), 3155-3164 (1990)
MEDLINE 91100426
REFERENCE 3 (bases 1 to 2156)
AUTHORS Korenberg, J.R., Chen, X.N., Tran, H. and Argraves, W.S.
TITLE Localization of the human gene for fibulin-1 (FBLN1) to chromosome
band 22q13.3
JOURNAL Cytogenet. Cell Genet. 68 (3-4), 192-193 (1995)
MEDLINE 95145011
FEATURES
source
Location/Qualifiers
1..2156
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2156
/evidence="experimental"
11..86
/sig_peptide
CDS 11..2062
/codon_start=1
/product="fibulin-1 C"
/db_xref="PID:g31419"
/db_xref="SWISS-PROT:P23144"
/translation="MERAAPSRVPLPILLGLGALLAAGVDADVLLEACCAADGHRMA
THQDCSLPYATESKECRMVQCCCHSLEELHCAATGTSILANEQDRCATPHGDNASLE
ATFVKRCHCLLGRAAQGGCEYSLMVGYCGQGVFRACCVKSOETGDLVDVGLQLE
DKIIIEVEEEDPYLNDRCRGKQCKQCRDTGDEVVSCFVGYQLLSDGVSCEDEVN
ECITGSHRGLGESINTVGSFRCDSSCGTGVELTDNSCKDIDECESGHNCILPD
FICONTLGSFRCPKLOCKSGFIODALNCIDINECLISAPCPIGHTCINTEGYTC
QKNVPCNGRGYHLNEEGRVDCVDECAPPAEPCGKGHRVNSPGSFRCFCKTGYFDG
ISRMVDVNECORYPGRCLGKHCNTLGLSCSVGFRLSVDRSCEDINBCSSPC
SQECANVYGSQCYRRGYQLSDVGVTCEDIDECALPTGGHICSYRCINIPGSFQCS
CPSSGYRLAPNRCODIDECVTGTHNCININETCFNIQAFRLAFCEPENRRSAAT
RCERLPCHENRECSKLPRIYHLSPPTNIQAPAVFMGMPSAVPDSVQSLATGG
NEEGFFTRKVSHPHSGVVALTKPVEPRDLLTVKMDLSRHGTVSFFAKLFIYFAE
L"
mat_peptide 87..2059
polyA_site 2156
BASE COUNT 420 a 654 c 647 g 435 t
ORIGIN
Query Match 2.1%; Score 33; DB 26; Length 2156;
Best Local Similarity 66.3%; Pred. No. 1.56e-04;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Db 1526 CAGTCAGCTGCCCTCGTCTGGCTACAGGCTGCCGCCCAATGGCCGAATGCCAAGAC 1585
QY 310 cagtgcctgtgtccatctcaggactccgctggcccaaatggaagagactgtctagat 369
Db 1586 ATTGATGAGTGTGATGGCATCCACAACTGCTCCATCAA 1626
QY 370 attgatgaatgtgcctctggttaaagtcatctgtccctacaa 410

```

```

RESULT 12
LOCUS HSF1BUA 2349 bp RNA PRI 21-AUG-1995
DEFINITION H.sapiens mRNA for fibulin-1 A.
ACCESSION X53741
NID g31414
KEYWORDS fibulin-1 A; glycoprotein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2349)
AUTHORS Argraves, W.S.
TITLE Direct Submission
JOURNAL Submitted (03-JUL-1990) Argraves W.S., American Red Cross, 15601
Crabbs Branch Way, Rockville, MD 20855, USA
REFERENCE 2 (bases 1 to 2349)
AUTHORS Argraves, W.S., Tran, H., Burgess, W.H. and Dickerson, K.
TITLE Fibulin is an extracellular matrix and plasma glycoprotein with
repeated domain structure
JOURNAL J. Cell Biol. 111 (6 Pt 2), 3155-3164 (1990)
MEDLINE 91100426
REFERENCE 3 (bases 1 to 2349)
AUTHORS Korenberg, J.R., Chen, X.N., Tran, H. and Argraves, W.S.
TITLE Localization of the human gene for fibulin-1 (FBLN1) to chromosome
band 22q13.3
JOURNAL Cytogenet. Cell Genet. 68 (3-4), 192-193 (1995)
MEDLINE 95145011
FEATURES
source
Location/Qualifiers
1..2349
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..2349
/tissue_type="placenta"
/chromosome="22q13.3"
1..2349
/evidence="experimental"
11..85
/sig_peptide
CDS 11..1711
/codon_start=1
/product="fibulin-1 A"
/db_xref="PID:g31415"
/db_xref="SWISS-PROT:P23142"
/translation="MERAAPSRVPLPILLGLGALLAAGVDADVLLEACCAADGHRMA
THQDCSLPYATESKECRMVQCCCHSLEELHCAATGTSILANEQDRCATPHGDNASLE
ATFVKRCHCLLGRAAQGGCEYSLMVGYCGQGVFRACCVKSOETGDLVDVGLQLE
DKIIIEVEEEDPYLNDRCRGKQCKQCRDTGDEVVSCFVGYQLLSDGVSCEDEVN
ECITGSHRGLGESINTVGSFRCDSSCGTGVELTDNSCKDIDECESGHNCILPD
FICONTLGSFRCPKLOCKSGFIODALNCIDINECLISAPCPIGHTCINTEGYTC
QKNVPCNGRGYHLNEEGRVDCVDECAPPAEPCGKGHRVNSPGSFRCFCKTGYFDG
ISRMVDVNECORYPGRCLGKHCNTLGLSCSVGFRLSVDRSCEDINBCSSPC
SQECANVYGSQCYRRGYQLSDVGVTCEDIDECALPTGGHICSYRCINIPGSFQCS
CPSSGYRLAPNRCODIDECVTGTHNCININETCFNIQAFRLAFCEPENRRSAAT
RCERLPCHENRECSKLPRIYHLSPPTNIQAPAVFMGMPSAVPDSVQSLATGG
NEEGFFTRKVSHPHSGVVALTKPVEPRDLLTVKMDLSRHGTVSFFAKLFIYFAE
L"
mat_peptide 86..1708
polyA_site 2349
BASE COUNT 520 a 668 c 682 g 479 t
ORIGIN
Query Match 2.1%; Score 33; DB 26; Length 2349;
Best Local Similarity 66.3%; Pred. No. 1.56e-04;
Matches 67; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
Db 1526 CAGTCAGCTGCCCTCGTCTGGCTACAGGCTGCCGCCCAATGGCCGAATGCCAAGAC 1585
QY 310 cagtgcctgtgtccatctcaggactccgctggcccaaatggaagagactgtctagat 369
Db 1586 ATTGATGAGTGTGATGGCATCCACAACTGCTCCATCAA 1626
QY 370 attgatgaatgtgcctctggttaaagtcatctgtccctacaa 410

```

```

RESULT 13
LOCUS U01244 2359 bp mRNA PRI 18-OCT-1996

```









```

RESULT 2
ID Q51746 standard; cDNA; 91 BP.
AC Q51746;
DT 31-MAY-1994 (first entry)
DE Oligonucleotide probe MK14-A
KW Oligonucleotide; DNA probe; mycobacteria; disease diagnosis;
OS .
KW Synthetic.
PN EP-571911-A.
PD 01-DEC-1993.
PF 24-MAY-1993; 108325.
PR 26-MAY-1992; US-889651.
PA (BECT ) BECTON DICKINSON CO.
PI Shank DD, Spears PA;
PR WPI; 93-378844/48.
PT New oligo:nucleotide probes specific for Mycobacteria - used for
PT detection and amplification of Mycobacteria nucleic acid in
PT samples
CC Claim 3; Page 14; 23pp; English.
CC Oligonucleotide probe MK14-A consists of nucleotides 5-95 of MK14
CC (Q51735). It hybridized to all spp. of mycobacteria tested, but
CC cross reacted to a few non-mycobacterial spp. The probe may
CC be useful as an initial screen for mycobacterial infection.
CC See also Q51735-45 and Q51747-59.
SQ Sequence 91 BP; 5 A; 17 C; 15 G; 4 T;

Query Match 2.6%; Score 41; DB 9; Length 91;
Best Local Similarity 5.6%; Pred. NO. 2.86e-10;
Matches 3; Conservative 47; Mismatches 3; Indels 1; Gaps 1;

Db 10 gssvhyvvhvshhshvhhvhhvsvvvhhvhhvhhvhhvhyhyvsvct 62
|:::: :::::: :::::: :::::: :::::: :::::: :::::: :::::: ::::::
Cp 981 gcgaatcactcgttcattcacctagggaacacacatctctgcagcgt 928

RESULT 3
ID N81164 standard; DNA; 204 BP.
AC N81164;
DT 08-NOV-1990 (first entry)
DE Base substituted E.coli beta-galactosidase alpha-fragment.
KW E.coli beta galactosidase alpha-fragment; base substitutions; ss.
OS Escherichia coli.
FH Key Location/Qualifiers
FT misc_feature 19..69 /*tag= a
FT /*function= multiple cloning site
FT 187..204 /*tag= b
FT primer_bind
FT PN EP-285123-A.
FT PD 05-MAY-1988.
FT PF 30-MAR-1988; 105163.
FT PR 03-APR-1987; US-034819.
FT PA (SUSO) SUOMEN SOKERI OY.
FT PI Lehtovaara P, Knowles J, Koivula A, Bamford J, Reinikainen T;
FT WPI: 88-279927/40.
PT Introducing random point mutations into nucleic acids -
PT by prepn of single stranded template, annealing a primer, elongation,
PT misincorporation, completion of molecules and screening.
PT Disclosure; P; English.
PS Random point mutations were introduced into the alpha fragment of
CC E.coli beta-galactosidase. The wild type sequence was obtained as a
CC single stranded template and an oligonucleotide was hybridised to
CC it to generate a popn of DNA molecules which terminate at all
CC possible nucleotide positions within a specified region. The
CC variable 3' ends generated in this way are used as primers for
CC reverse transcriptase. Nucleotides are misincorporated by the
CC transcriptase and the molecules are completed to forms that can be
CC amplified and then expressed in a suitable host-vector system.
CC The sequence covers all 176 diff base substitutions, most of which
CC occurred singularly in any given mutant.
CC See also P80575.

```

```

KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /*note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or comsps. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed activity
CC allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.1%; Score 34; DB 12; Length 114;
Best Local Similarity 1.9%; Pred. No. 3.70e-06;
Matches 2; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnn 62
Cp 940 cctcgcaggctcgtctctatgtcattcttcacggcttctctctctctctcat 881
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 108
Cp 880 cctcagccctctctctctctctctctctctctctctctctctctctcc 835

RESULT 6
ID Q70468 standard; DNA; 114 BP.
AC Q70468;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /*note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYN-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI; 94-279739/34.
DR P-PSDB; R65153.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)16(TGC)(NNB)12(NNB)16(TGC)(NNB)10Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in Q70466-68.
CC Other specific peptides generated by these generic sequences are shown in
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC peptide between the 2 domains. The oligonucleotides are also designed so
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC or comsps. comprising a TSAR binding domain can be used in vivo to
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC radioisotope, peptide, toxin or enzyme, to the specific target or on the
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.1%; Score 34; DB 12; Length 114;
Best Local Similarity 1.9%; Pred. No. 3.70e-06;
Matches 2; Conservative 33; Mismatches 71; Indels 0; Gaps 0;

Db 3 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnnn 62
Cp 940 cctcgcaggctcgtctctatgtcattcttcacggcttctctctctctctcat 881
Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 108
Cp 880 cctcagccctctctctctctctctctctctctctctctctctctctcc 835

RESULT 7
ID Q70465 standard; DNA; 114 BP.
AC Q70465;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /*note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994. U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.

```





```

FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65150 and R65151.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)6(TGC)(NNB)12(NNB)14(TGC)(NNB)3Y. X
CC and Y are flanking restriction sites (X is not the same as Y) that are
CC not specified further. Other generic sequences are shown in
CC Q70465.
CC R65151-54. TSARs are concatenated heterofunctional proteins or peptides,
CC comprising at least two functional regions - a binding domain with
CC affinity for a ligand and a second effector peptide portion that is
CC chemically or biologically active. They may further comprise a linker
CC that the expressed peptide contains 2 or 4 cysteine residues positioned
CC in, or flanking, the unpredicted or variant residues. These residues
CC confer some degree of conformational rigidity to the peptides. The TSARs
CC deliver a chemically or biologically active moiety, eg. metal ion,
CC cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need
CC for complex methods of hybridoma formation or in vivo antibody
CC production. The TSARs are easily characterised and have designed
CC activity allowing direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 2 C; 2 G; 2 T;

Query Match 2.0%; Score 32; DB 12; Length 114;
Best Local Similarity 3.6%; Pred. No. 4.97e-05;
Matches 4; Conservative 32; Mismatches 76; Indels 0; Gaps 0;

Dn 3 bnnbnnbnnbnnbgtcnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 62
1038 taticgttgactgcagcttcaatcgtggtatgctgaggaacaggtatagaaga 1097

Db 63 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnb 114
Qy 1098 tgatttgactgaatcctgcgtgatagatagctatgcttcttatg 1149

RESULT 13
ID Q70466 standard; DNA; 114 BP.
AC Q70466;
DT 05-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR-9 peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "this sequence represents 'Z'; Z can be a
FT sequence of 6, 9 or 12 nucleotides (see
FT comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R58378.
PT Identifying proteins or peptide(s) which bind a ligand - by

```

```

PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R65152.
PT Identifying proteins or peptide(s) which bind a ligand - by
PT screening a recombinant vector library expressing fusion proteins
PT comprising a binding domain and an effector domain
PS Disclosure; Page 35; 255pp; English.
CC Q70466 is a generic DNA sequence used to generate random TSAR (Totally
CC Synthetic Affinity Reagents) peptides. This generic formula can also be
CC represented as follows: X(NNB)1(TGC)(NNB)10(TGC)2(NNB)42(NNB)8(TGC)(NNB)
CC -9Y. X and Y are flanking restriction sites (X is not the same as Y)
CC that are not specified further. Other generic sequences are shown in
CC Q70466-68. Other specific peptides generated by these generic sequences
CC are shown in R65151-54. TSARs are concatenated heterofunctional proteins
CC or peptides, comprising at least two functional regions - a binding
CC domain with affinity for a ligand and a second effector peptide portion
CC that is chemically or biologically active. They may further comprise a
CC linker peptide between the 2 domains. The oligonucleotides are also
CC designed so that the expressed peptide contains 2 or 4 cysteine residues
CC positioned in, or flanking, the unpredicted or variant residues. These
CC residues confer some degree of conformational rigidity to the peptides.
CC The TSARs or compsns, comprising a TSAR binding domain can be used in
CC vivo to deliver a chemically or biologically active moiety, eg. metal
CC ion, radioisotope, peptide, toxin or enzyme, to the specific target or
CC on the cell. They can also replace the function of macromolecules, eg.
CC monoclonal or polyclonal antibodies and therefore circumvent the need for
CC complex methods of hybridoma formation or in vivo antibody production.
CC The TSARs are easily characterised and have designed activity allowing
CC direct and rapid detection in a screening process.
SQ Sequence 114 BP; 0 A; 4 C; 4 G; 4 T;

Query Match 2.0%; Score 32; DB 12; Length 114;
Best Local Similarity 5.7%; Pred. No. 4.97e-05;
Matches 6; Conservative 30; Mismatches 70; Indels 0; Gaps 0;

Db 9 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnnn 68
Cp 997 ttggaccagaatcaggcggaatcacctgtcttcacctcagggaacacacatctc 938

Db 69 bnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnnbnn 114
Cp 937 ctgcaggcttcgctctctctatgctattcttcagggtctctctc 892

RESULT 14
ID Q70470 standard; DNA; 114 BP.
AC Q70470;
DT 10-APR-1995 (first entry)
DE Generic DNA sequence to generate a random TSAR peptide library.
KW TSAR; totally synthetic affinity reagent; synthetic; binding domain;
KW effector domain; concatenated heterofunctional protein; linker;
KW direct; rapid; detection; screening; treatment; generic; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 55..60
FT /*tag= a
FT /note= "encoded by Z (see comments)"
PN WO9418318-A.
PD 18-AUG-1994.
PF 01-FEB-1994; U00977.
PR 01-FEB-1993; US-013416.
PR 30-DEC-1993; US-176500.
PR 31-JAN-1994; US-189331.
PA (UYNC-) UNIV NORTH CAROLINA.
PI Fowlkes DM, Kay BK;
DR WPI: 94-279739/34.
DR P-PSDB; R58378.
PT Identifying proteins or peptide(s) which bind a ligand - by

```





\*\*\*\*\*  
W P S R L H  
(TM)  
\*\*\*\*\*

Release 3.1A John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993-1998 University of Edinburgh, U.K.  
Distribution rights by Oxford Molecular Ltd

srch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm  
Run on: Sat Apr 24 22:35:35 1999; MasPar time 2350.24 Seconds  
Tabular output not generated. 1227.509 Million cell updates/sec

Title: >US-08-968-800-2  
Description: (1-1611) from US08968800.seq  
Perfect Score: 1607  
N.A. Sequence: 1 gctggagaagaacagcaa.....aattactagctgaaaaatg 1611  
Comp: ccgaccctttcttctgtcgtt.....ttaagatgacgaatttttaac

Scoring table: TABLE default  
Gap 6  
Nmatch STD : Dbase 0; Query 0

Searched: 2275026 seqs, 895388244 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: embl-est56  
Database: genbank-est109  
5:gb\_est1 6:gb\_est10 7:gb\_est11 8:gb\_est12 9:gb\_est13  
10:gb\_est14 11:gb\_est15 12:gb\_est16 13:gb\_est17  
14:gb\_est18 15:gb\_est19 16:gb\_est2 17:gb\_est20  
18:gb\_est21 19:gb\_est3 20:gb\_est4 21:gb\_est5 22:gb\_est6  
23:gb\_est7 24:gb\_est8 25:gb\_est9 26:gb\_gss1 27:gb\_gss2  
28:gb\_gss3 29:gb\_gss4

Statistics: Mean 11.687; Variance 2.100; scale 5.565

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	555	34.5	626	AA195615	zr37b10.r1 Soares NHM	0.00e+00
2	451	28.1	508	AA025649	ze85d03.r1 Soares feta	0.00e+00
3	438	27.3	472	AA436507	zv08c07.r1 Soares NHM	0.00e+00
4	398	24.8	446	R99817	yq69b11.r1 Homo sapien	0.00e+00
5	355	22.1	473	R76984	yf66d08.r1 Homo sapien	0.00e+00
6	333	22.0	411	R64051	y122c04.r1 Homo sapien	0.00e+00
7	334	22.0	454	H03336	y744a05.r1 Homo sapien	0.00e+00
8	332	20.7	429	AA009900	ze83d06.r1 Soares feta	0.00e+00
9	331	20.6	341	R68291	y106a03.r1 Homo sapien	0.00e+00
10	311	19.4	442	T65608	yc76b01.r1 Homo sapien	0.00e+00
11	286	18.4	406	AI057064	oz07b04.r1 Soares_feta	0.00e+00
12	285	17.7	344	W80921	zd90b04.r1 Soares_feta	0.00e+00
13	277	17.2	347	R11377	yf42e08.r1 Homo sapien	0.00e+00

14	252	15.7	374	5	R27678	yh64e08.r1 Homo sapien	0.00e+00
15	249	15.5	364	5	R27153	yh46e02.r1 Homo sapien	0.00e+00
16	247	15.4	256	19	HSC32G011	H. sapiens partial CDN	0.00e+00
17	248	15.4	325	5	R24466	yh48c12.r1 Homo sapien	0.00e+00
18	236	14.7	329	5	R27258	yh53d04.r1 Homo sapien	0.00e+00
19	234	14.6	374	5	R24180	yh29g10.r1 Homo sapien	0.00e+00
20	218	13.6	270	5	R32182	yh67f04.r1 Homo sapien	0.00e+00
21	197	12.3	388	16	R64033	y119h01.r1 Homo sapien	0.00e+00
22	197	12.3	215	6	C17012	Human placenta cDNA 5'	0.00e+00
23	186	11.6	353	5	R34484	yh65f01.r1 Homo sapien	0.00e+00
24	178	11.1	213	19	N39776	yh65a12.r1 Homo sapien	0.00e+00
25	137	8.5	527	11	AA707654	z129e09.s1 Soares_feta	9.47e-230
26	118	7.3	420	18	AA913032	o134a06.s1 Soares_NFL	5.66e-188
27	114	7.1	506	18	AI057014	oz06d10.x1 Soares_feta	2.90e-179
28	96	6.0	474	17	AI023885	ow69c06.s1 Soares_feta	1.58e-140
29	86	5.4	485	22	AA009486	ze83d06.s1 Soares_feta	2.06e-119
30	77	4.8	471	21	AA025548	ze85d03.s1 Soares_feta	1.00e-100
31	64	4.0	252	12	AA754459	97SN1787 Rice Immature	1.96e-74
32	61	3.8	252	12	AA754459	97SN1787 Rice Immature	1.64e-68
33	61	3.8	457	23	AA259930	vab3c02.r1 Soares_mous	1.49e-66
34	60	3.7	439	17	AI022137	ow68e06.x1 Soares_feta	1.49e-66
35	43	2.7	247	12	AA754458	97SN1784 Rice Immature	1.09e-34
36	44	2.7	247	12	AA754458	97SN1784 Rice Immature	1.09e-36
37	44	2.7	423	11	AA702522	z180d04.s1 Soares_feta	1.80e-36
38	39	2.4	334	10	AA614649	np48a02.s1 NCI-IGAP.Br	1.01e-27
39	35	2.2	2275	11	AF034173	Homo sapiens ntcon2 co	4.84e-21
40	33	2.1	360	6	C18091	Human placenta cDNA 5'	7.85e-18
41	33	2.1	408	16	R69676	y14f01.r1 Homo sapien	7.85e-18
42	33	2.1	420	16	R71008	y150g07.r1 Homo sapien	7.85e-18
43	34	2.1	452	6	AA130066	z133c11.r1 Soares_preg	2.00e-19
44	33	2.1	467	16	R66735	y133b06.r1 Homo sapien	7.85e-18
45	33	2.1	479	16	H01614	yj21d02.r1 Homo sapien	7.85e-18

ALIGNMENTS

1  
RESULT AA195615 626 bp mRNA EST 19-MAY-1997  
LOCUS zr37b10.r1 Soares NHMPU S1 Homo sapiens cDNA clone 665563 5', mRNA  
DEFINITION sequence.  
ACCESSION AA195615  
NID g1783792  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;  
Homo.  
REFERENCE 1 (bases 1 to 626)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.  
WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1094 Std Error: 0.00  
High quality sequence stop: 477.  
Location/Qualifiers  
1. 626  
/organism="Homo sapiens"  
/note="Organ: mixed (see below); Vector: p7T3b-Pac (Pharmacia) with a modified polylinker; Site\_1: I; Site\_2: Eco RI; Equal amounts of plasmid DNA from three

FEATURES

source



```

FEATURES
source
Location/Qualifiers
1. .472
/organism="Homo sapiens"
/note="Organ: mixed (see below); Vector: p7T3D-Pac
(Pharmacia) with a modified polylinker; Site 1: Not I;
Site 2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NBHM, pregnant uterus
NbHPu, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
/db_xref="taxon:9606"
/clone="753036"
/clone_lib="Soares NbHPu S1"
/tissue_type="pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
<1. .>472
/db_xref="GDB:5975877"
138 a 139 g 122 t 1 others
mRNA
BASE COUNT

```

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu

High quality sequence stops: 326  
 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.

#### FEATURES

Location/Qualifiers

1..446

/organism="Homo sapiens"

/clone="201021"

<1..>446

#### BASE COUNT

138 a 93 c 104 g 107 t 4 others

#### ORIGIN

Query Match 24.8%; Score 398; DB 16; Length 446;

Best Local Similarity 97.1%; Pred. No. 0.00e+00;

Matches 433; Conservative 0; Mismatches 8; Indels 5; Gaps 5;

1 ACAGAGAGAGGCGCACAGTGTGTCATCCCTCAGGACTCCGCTGGCCCAATGGA 60

QY 295 acagaaagagggccacagtgctgtccatccctcagactccgctgcccacaaatgga 354

Db 61 AGAGACTGTCTAGATATTGATGTAATGTGCTGCTGTAAGTCATCTGTCCTACATCGA 120

QY 355 agagactgtctagatattgatgtgctgtccatccctcagactccgctgcccacaaatgga 414

Db 121 AGATGTGTGAACACATTTGGAAGTACTTACTGCAAAATGCAATGCTGTTTGAAGTCAAA 180

QY 415 agatgtgtgaacacatttggagctactactgcaaatgcaattgctggaactgcaa 474

Db 181 TATATCAGTGGACGATATGACTGTATAGATATAAATGATGATGATGATGATGATGATGAT 240

QY 475 tatatcagtgagatgactgtatagataataatgaatgaatgaatgaatgaatgaatgaatgaat 534

Db 241 TGCAGCACCATGCCAATTCCTCAATFACCCAGGGTCCCTCAAGTGTAAATGCAAGCAG 300

QY 535 tgcagccacatgccaatgtcttcaatcaccacagggctcctcaagttaaatgaagcag 594

Db 301 GGATATAAGGCAATGGACTTCGGTGTCTGCTATCCCTGGAAATTCGTGGAAGGAG 360

QY 595 ggataaaaggcaatggacttggctgtctgtctatccctg-aaattctctg-gaaggag 652

Db 361 TCCTCAGACACCTGGGTACCNFCAAGCAGGATCAAGGAGTGTCTGTTTACACAAA 420

QY 653 tctcagagcactgg-taccatcaagacagaaat-caagaagtgtctgtct-cacaaaa 709

421 ACAGCATGANAAGNAGGCAAAATT 446

QY 710 acagcatgaaaaaaggaagcacaatt 735

#### RESULT

##### LOCUS

R76984 473 bp mRNA EST 06-JUN-1995

##### DEFINITION

vi66d08.r1 Homo sapiens CDNA clone 144207 5'

##### ACCESSION

R76984

9851616

EST.

human clone-144207 library=Soares placenta Nb2HP vector=PT73D

(Pharmacia) with a modified polylinker host=DH10B (ampicillin

resistant) primer=M13RPI Rsite1-Not I Rsite2-Eco RI Female placenta

obtained at birth (full term). 1st strand cDNA was primed with a

Not I - oligo(dT) primer [5,

AACTGAAGAGATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3'], double-stranded

cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not

I and cloned into the Not I and Eco RI sites of the modified pT73

vector. Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo.

Homo sapiens

Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;

#### REFERENCE

##### AUTHORS

Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 473)  
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,  
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,  
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and  
 Wilson, R.

#### TITLE

##### JOURNAL

##### COMMENT

1..473

/organism="Homo sapiens"

/clone="144207"

132 a 78 c 106 g 147 t 10 others

Query Match 22.1%; Score 355; DB 16; Length 473;

Best Local Similarity 96.8%; Pred. No. 0.00e+00;

Matches 365; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

Db 1 GTTTCCTCTTGTATNACCGCTGGCCANGNACAAAGTCGGAAACTTCGAGTGTGTGCA 60

QY 1235 gtttgcctttgtattaccgctgctggcgagacaaagtgcggaacttcgagtggttggga 1294

Db 61 AAACAGTAACAATGCCCTGG-ATGGGAGAGACACACAGTGAGGATGAAAGTGAAGA 119

QY 1295 aaacagcaacaatgccctggcatgggagagaccacagtgaggatgaaagtgaaga 1354

Db 120 CAGGAAATTCAGTGTGTATCAAGGAAGTGTATGCTACCAAGACATCAATTTTGAAGCAG 179

QY 1355 cagggaattcagttgtatcaaggactgtctaccacaaagcatcttttgaagcag 1414

Db 180 AACGTGCAAGGGCAAAACCGCGAAATCGCAGTGGGCTCTTGTCTTTCAGGCT 239

QY 1415 aacgtggcaaggcgaacacccggcgaatcgagtgagtggtctgtcttcaggct 1474

Db 240 TATGTCAGATAGCCTTTTATCTGTGGTGAATGAAATGTTTACTATCTTATATTTGACTT 299

QY 1475 tatgtccagatagccttttatctgtggannnctgaatgggtactatctttatattgactt 1534

Db 300 TGTATGTCAGTCTCCTGGTTTTTTTGTATTTGATATGATCATAGACCTCGCAATTTAGAAT 359

QY 1535 tgtatgtcagttccctgggtttttttatattgcatcatagacacctggcatttttaaat 1594

Db 360 TACTAGCTGAAAANTTTG-376

QY 1595 tactagtgaataattg 1611

#### RESULT

##### LOCUS

R64051 411 bp mRNA EST

##### DEFINITION

Y122C04.r1 Homo sapiens CDNA clone 139974 5'

##### ACCESSION

R64051

9835930

EST.

human clone=139974 library=Soares placenta Nb2HP vector=PT73D

(Pharmacia) with a modified polylinker host=DH10B (ampicillin

resistant) primer=M13RPI Rsite1-Not I Rsite2-Eco RI Female placenta

obtained at birth (full term). 1st strand cDNA was primed with a

Not I - oligo(dT) primer [5,

AACTGGAAGAAATTCGGCGCCGAGGAATTTTTTTTTTTTTTTT 3'], double-stranded

CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 411)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 314  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source  
1. 411  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="139974"

BASE COUNT 121 a 80 c 107 g 100 t 3 others  
ORIGIN

Query Match 22.0%; Score 353; DB 16; Length 411;  
Best Local Similarity 95.8%; Pred. No. 0.00e+00;  
Matches 392; Conservative 0; Mismatches 12; Indels 5; Gaps 5;

Db 2 CAAGGAAGCGCTAACCTCCAACTGGAACATAAGATTAAATCTCGGTGACTGC 61  
QY 994 caaaggaagcgctaactccaaactggaacacaggaatttaataatctcgttgactgc 1053  
Db 62 AGCTTCATCGGATCTGACATGGAACAGGATAGAGACATGATTGACTGGAAT 121  
QY 1054 agcttcaatcgtggtatgactggaacacaggaatttaataatctcgttgactggaat 1113  
Db 122 CTGCTGATCGAGATAATGCTATTGGCTTCTATATGGCAGTTCGGCTTGGCAGGTCAC 181  
QY 1114 cctgctgacgagataaagtctattggtctctatatgagcagttccggtgaggtcagc 1173  
Db 182 AAGAAGACATTGGCGGATGGAACCTTCTACCTGACCTGCAACCCCAAGCAACTTC 241  
QY 1174 atgaagacattggtggtggaactctcctacctgacctgcaaccccaagcaacttc 1233  
Db 242 TGTTCCTCTTTGATTACCGCTGGCGGAGGACAAAGTCGGGAACTTCGAGTGTGT 301  
QY 1234 tgttgcctcttgattacggtggtggtggtggtggtggtggtggtggtggtggtggt 1292  
Db 302 GAAACACAGTACATGCTCGTATGCTGATGGGAGAACACACGAGTGAGGTTGAAAGTGGG 361  
QY 1293 gaaacacagtaacatacctcctggtcagtggtggtggtggtggtggtggtggtggtggt 1350  
Db 362 AAGNAGGAGGAAATTCAGTCTTTTCAAGGACCTGNTGTTACCCCAAG 410  
QY 1351 aagacagggaaaatt-cagttgtat-caaggaaactgacttaccacaaaag 1397

RESULT 7 H03936 454 bp mRNA EST 20-JUN-1995  
LOCUS Y144a05.s1 Homo sapiens CDNA clone 151568 3'  
DEFINITION H03936  
ACCESSION

9866869

KEYWORDS  
SOURCE

EST.  
human clone=151568 library=Soares placenta Nb2HP vector=pY7r3D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=Promega -2lm13 Rsite1-Not I Rsite2=Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - Oligo(dT) primer [5' AACGGAAGATTCGCGCGCAGGAATTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pYT3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens  
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 454)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE

The WashU-Merck EST Project

JOURNAL

Unpublished (1995)

COMMENT

Contact: Wilson RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 354  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: polyt not found.

FEATURES  
source

1. 454  
Location/Qualifiers  
/organism="Homo sapiens"  
/clone="151568"

BASE COUNT 144 a 86 c 119 g 102 t 3 others  
ORIGIN

Query Match 22.0%; Score 354; DB 16; Length 454;  
Best Local Similarity 97.1%; Pred. No. 0.00e+00;  
Matches 405; Conservative 0; Mismatches 4; Indels 8; Gaps 6;

Db 39 AGCGGGAACTCTCATGGAGGTAAAAAGGAATGAAGAGAAATGAAGAGGGCTTGA 98  
QY 819 agcgggaaactctcatggaggttaaaaggaatgaagagaaatgaagagggcttga 878  
Db 99 GGATGAGAAAGAGAGAGAAAGCCCTGAAGAATGACATAGAGAGCGAANC-TGGCAGG 157  
QY 879 ggaatgagaaagagagagaaagccctgaagaatgacatagagagagcagctgcgag 938  
Db 158 AGATGTGTTTTCCCTTAAGTGAATGAAGCAGGTGAATTCGGCTGATTCGTGCTCAAG 217  
QY 939 agatgtgttttcccttaagtgaatgaagcaggtgaatcgctgattctggtctcaag 998  
Db 218 GAAAGCGCTAACCTCCAACTGGAACATAAGCAGATTAAATATCTCGGTTGACTGCAG 277  
QY 999 gaaagcgtcaactccaaactggaacataaag---atttaaatatctcgttgactgcag 1055  
Db 278 CTTCAATCATGGGATCTGTGACTGGAAACAGGATAGAGAAGATGATTTTTCACCTGGAATCC 337  
QY 1056 cttcaatcatgggattctgtgactggaaacaggatagagagatgattttgactgcag 1115  
Db 338 TGCTGATCGAGATAATGCTATTGCTTCTATATGGCAGTTCGCCCTTCGCGAGTTTAC 397  
QY 1116 tgctgatcgagataatgctattgcttcttatatggc-agttccggccttgaggt-cac 1173

Db 398 AAGAAGACATGGCCGCTTTGAACCTCTCTACCTGACCTGGCAACCCCAAGCA 454  
 QY 1174 atgaagacattggcc-gattgaacttctctactacgtgacctg-caaccccaagca 1228

RESULT 8  
 LOCUS AA009900 429 bp mRNA EST 01-FEB-1997  
 DEFINITION ze83d06.r1 Soares fetal heart NBHH19W Homo sapiens cDNA clone  
 365579 5', mRNA sequence.  
 ACCESSION AA009900  
 NID 91470927  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
 WashU-Merck EST Project  
 Unpublished (1995)

TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: mob.REGA+RT  
 High quality sequence stop: 372.  
 Location/Qualifiers  
 1..429  
 /organism="Homo sapiens"  
 /note="Organ: heart; Vector: pTT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCAGAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bernaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."  
 /db\_xref="taxon:9606"  
 /clone="365579"  
 /clone\_lib="Soares fetal heart NBHH19W"  
 /sex="unknown"  
 /dev\_stage="19 weeks"  
 /lab\_host="DH10B (ampicillin resistant)"  
 <1..>429

mRNA  
 BASE COUNT 135 a 76 c 117 g 96 t 5 others  
 ORIGIN

Query Match 20.7%; Score 332; DB 22; Length 429;  
 Best Local Similarity 96.8%; Pred. No. 0.00e+00;  
 Matches 399; Conservative 0; Mismatches 2; Indels 11; Gaps 11;

Db 1 GAGATAGTTCCAGAGCGGCACTCTCATGGAGTAAAGGGAATGAAGAAATG 60  
 QY 805 gagatagttccagagcgcggaactctctgagggtataaaaggaatgaagaaatg 864  
 Db 61 AAAGAGGGCTTGAGGATGAGAAAGAGAGAAAGCCCTGAAGATGACATAGAGAG 120  
 QY 865 aaagagggcttgaggatgagaaagagagaaagccctggaatgacatagaggag 924

Db 121 CGAA-CCTGCGAGGAGATGTGTTTTCCCTAAGGTGAATGAAGCAGGTGAATTCGCCCTG 179  
 QY 925 cgaagcctgcgagagatgtgttttccctaaagtgaatgaagcagggaattgcgcctg 984

Db 180 ATTTCGGTCCAAAGGAAAGCGCTAACTTCCAACTGGAACATGAAGATTTTAAATATCTC 239  
 QY 985 attct-ggttcaaaaggaagcgtaacttccaaactggaacataaagatttaataatctc 1043

Db 240 GGTGACTGCAGCTTCAATCATNGGATCTGTGACTGGAAACAGGATAGAGGAAGATGAT 299  
 QY 1044 ggttgactgcagcttcaatcat-gggaatctgtgactggaacagagataga-gaagatgat 1101

Db 300 TTGACTGGAATCTGCTGNTGCGAGATAATGCTATTNGGCTTCTATATGCGAGTTCGCCG 359  
 QY 1102 ttgactggaatcctgctgacgataatgctatt-ggcttctatgagcagttcc-gg 1159

Db 360 CTTTGGCAGGTCCACAAGAAAGACATTTGGCCCGGATTTGAAACTTCTCCTA 411  
 QY 1160 cttt-ggcaggtc-acatgaagacatt-ggcc-gatt-gaaacttctccta 1206

RESULT 9  
 LOCUS R68291 341 bp mRNA EST 01-JUN-1995  
 DEFINITION Y106a03.r1 Homo sapiens cDNA clone 138412 5'.  
 ACCESSION R68291  
 NID 9841808  
 KEYWORDS EST.  
 SOURCE human clone=138412 library=Soares placenta Nb2HP vector-pT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) primer=M13RP1 Rsite1-Not I Rsite2-Eco RI Female placenta obtained at birth (full term). 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' ACTGGAAGATTCGCGCCGAGGATTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bernaldo.

ORGANISM  
 Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.  
 REFERENCE 1 (bases 1 to 341)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
 The WashU-Merck EST Project  
 Unpublished (1995)

TITLE  
 JOURNAL  
 COMMENT

Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 292  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Location/Qualifiers  
 1..341  
 /organism="Homo sapiens"  
 /clone="138412"

BASE COUNT 95 a 60 c 87 g 98 t 1 others  
 ORIGIN

Query Match 20.6%; Score 331; DB 16; Length 341;  
 Best Local Similarity 97.9%; Pred. No. 0.00e+00;  
 Matches 334; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Db 1 TACCGCTGGCCGANGACAAAGTCGGGAAACTTCGAGTGTGTGTAAGAAACAGTAAACAT 60  
 QY 1249 taccgctggccgagacaaagtcgggaaacttcgagtggttgtaaaacagtaacaat 1308  
 Db 61 GCGCTGCATGGGAGACACACAGTGCAGATGAAGAGTGAAGACAGGGAATTCAG 120  
 QY 1309 gccctggcagtggaagaccacagtgagatgaaagtggagacaggggaaattcag 1368  
 Db 121 TTGTATCAAGAACTGATGCTACCAAAAAGCATCATTTTGAAGCAGAGAGTGGCAAGGCG 180  
 QY 1369 ttgtatcaagaaactgatgtaccacaaagcatcatttttgaagcagaacgtggcaagggc 1428  
 Db 181 AAAACCGCGGAAATCGCAGTGGATGGCTGTGCTGTGTTTACAGGTTTATGTCCAGATAGC 240  
 QY 1429 aaaacccggcaaatcgcaagtggcgtctgttctgttcttcagcttatgtccagatagc 1488  
 Db 241 CTTTATCTGGGATGACTGATGTACTATCTATCTTTATATTTGATGTGATGTCTAGTCC 300  
 1489 ctttattctgtggannnctgaatgactatcttatttattgactttgtatgtcagttcc 1548  
 Db 301 CTGTTTTTTTGTATTGATCATCATAGGACCTCTGTCTATTTT 341  
 QY 1549 ctggtttttttgatattgcatcatagggacctctctggtcatttt 1589

RESULT 10  
 LOCUS T65608 442 bp mRNA EST  
 DEFINITION YC76B01.r1 Homo sapiens CDNA clone 21599 5' 07-MAR-1995  
 ACCESSION T65608  
 NID 9574653  
 KEYWORDS  
 SOURCE human clone-21599 library-Soares infant brain INIB vector-Lafmid BA host=DH10B (ampicillin resistant) primer=M13RPI Rsite1-Not I Rsite2-Hind III Whole brain from a 73 days post natal female. 1st strand CDNA was primed with a Not I - oligo(dT) primer [5' AACTGGAAGATTCGGCGCGAGAGATTTTTTTTTTTT 3']; double-stranded CDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lafmid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.

ORGANISM Homo sapiens  
 Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 442)  
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.  
 TITLE The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT GDB: G00-393-946  
 Contact: Wilson RK  
 WashU-Merck EST Project  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@watson.wustl.edu  
 High quality sequence stops: 305  
 Source: IMAGE Consortium, LLNL  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
 source  
 1..442  
 /organism="Homo sapiens"  
 /clone="21599"  
 BASE COUNT 144 a 81 c 115 g 94 t 8 others  
 ORIGIN  
 Query Match 19.4%; Score 311; DB 5; Length 442;

Best Local Similarity 95.8%; Pred. No. 0.00e+00;  
 Matches 345; Conservative 0; Mismatches 11; Indels 4; Gaps 4;  
 Db 6 TATAAGGCAATGGACTTCGGTGTCTGCTATCCCTGAAATCTCTGAGGAAGCCCTC 65  
 QY 598 tataaggcaatggacttcggtgtctgtctatccctgaaattctgtgaagaagctc 657  
 Db 66 AGAGACCTCGGACCATCAAGACAGAAATCAAGAACTTGCTTCTCACAACACAGCATG 125  
 QY 658 agagacctggtaccatcaaacagaaatcaagaagtctgtctcacaacacagcatg 717  
 Db 126 AAAAGAAGCGCAAAATTAATAATTTACCCAGAACCCACAGGNTCTACCCCTAAG 185  
 QY 718 aaaagaagcgcaaaattaaatgttaccacagaccaccagagactctaccctaaag 777  
 Db 186 GTGAACTTCGACCCCTCAACTATGAAGAGATAGTTTCCAGAGCGCGGACACTCTCATGA 245  
 QY 778 gtgaacttcgaccttcactatgaagagatagtttccagagcgaggacctctcatga 837  
 Db 246 GGTAAAAAGGGAATGAAGAGAAATGAAGAGGGGCTTGAGGATGAGAAAGAGAGAG 305  
 QY 838 ggtaaaaagggaatgaagagaaatgaagaggggcttgagagatgagaaagagaag 897  
 Db 306 AAAGCCCTGAAGGATTGACATAGGAGGCGACCTTTTCGAGGGAGATGTTTTTTTCC 365  
 QY 898 aaagccctgaagaat-gacatag-agg-agcgaagcctgcgagg-agatgtgttttccc 953

RESULT 11  
 LOCUS AI057064 406 bp mRNA EST  
 DEFINITION Oz07b04.x1 Soares fetal\_liver\_spleen\_lNFLS\_S1 Homo sapiens CDNA clone IMAGE:1674607 3', mRNA sequence.  
 ACCESSION AI057064  
 NID 93330853  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 406)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
 Insert Length: 757 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 372.

FEATURES  
 Location/Qualifiers  
 1..406  
 /organism="Homo sapiens"  
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Pac I; Site\_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand CDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAAATTAAGATCTTTTTTTTTTTTTTTTTTTT 3'], double-stranded CDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1674607"  
 /clone\_lib="Soares\_fetal\_liver\_spleen\_lNFLS\_S1"  
 /sex="male"  
 /dev\_stage="20 week-post conception fetus"  
 /lab\_host="DH10B (ampicillin resistant)"





This clone is available royalty-free through LLNL; contact the SOURCE: Image Consortium, LLC for further information. IMAGE Consortium (info@image.llnl.gov) for further information.

### Location/Qualifiers

Genotype	96 a	69 c	96 q	83 t	3 others
/organism="Homo sapiens"					
/clone="129542"					

```
Query Match      15.7%; Score 252; DB 5; Length 374;
Best Local Similarity 95.9%; Pred. No. 0.00e+00;
Matches 280: Conservative 0; Mismatches 8; Indels 4; Gaps 3;
```

1 GGTGACTGCAGCTTCAATCATGGATCTGTGACTGGAAACAGGATAGAGAAGATGATT 60

1044 qgttgactgcagcttcaatcatgggatatctgtgactggaacacaggatagagaaqatgattt 1103

61 TGACTGGAATCCTGCTGATCGAGATAATGCTATTGGCTTCTATATGGCAGTTCGGCCTT 120

1104 tgactggaatcctgctgacgagataaatgctattggccttctatatggcagattccggcctt 1163

121 GGCAGGTCACAAGAAAGACATTGGCCGATTGAAACTTCTCCTACCTGACCTGCAACCCCA 180

1164 qqcaqgtcacatqaaaqacattqcccqattqaaacttctcctacctqacctqcaaaccca 1223

181 AAGCAACTTCTGTTTGCTCTTTGATTACCGGCTGGCNCAGGACAAAGTCGGGAAACTTC 240

1224 aaqcaacttctgttttgctcttttgattaccqgctqccqgaacaaagtctggg-aaacttc 1282

241 GAGTGTTTGTGAAAAACAGTAACAATGCCCTGGGCNTGGGGAGAAGACCACGAGTGAGGG 300

1283 qagttatttttqtgaaaaacagttaacaatatgccctgg-cattggg-aggaaagaccacgaagtgaag- 1339

301 ATGAAAAGTTGGAAGACCCAGGGGAAATTCAGTTGTAT 338

1340 atgaaaaagt-qqaaagac-aggqaaaaattc-agtttatat 1374

LOCUS	EST	MRNA	374 bp	25-APR-1995
SULT1A4				
LOCUS R27678				

DEFINITION Yn64e08.r1 Homo sapiens CDNA clone 134534 5'.

LD  
9/83813  
KEYWORDS  
EST.

SOURCE  
human clone=I34534 Library=soares placenta NB2HP vector=PT/T3D  
(Pharmacia) with a modified polylinker host=DH10B (ampicillin

resistant) primer=MI3RPI Rsite1=Not I Rsite2=Eco RI Female plac obtained at birth (full term). 1st strand cDNA was primed with

Not I - oligo(dT) primer (5',  
AACTGGAAAGAAATTCGGGGCCGACGGAATTTT  
3', double-stranded)

CDNA was ligated to Eco RI adaptors (Pharmacia), digested with T and cloned into the Not I and Eco RI sites of the modified pT

vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Ronaldo

ORGANISM Homo sapiens  
Eucaryotae: Metazoa: Chordata: Vertebrata: Gnathostomata: Mammalia

Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 271)

**AUTHORS**  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,  
Holman M., Hultman M., Kucaba T., Lannon C., Marra M.

Parsons, J., Rifkin, L., Rohlffing, T., Soares, M., Tan, F.,

Wilson, R.

Contact: Wilton RK  
WashU-Merck EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
High quality sequence stops: 283  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES  
Location/Qualifiers  
1..364

Source  
/organism="Homo sapiens"  
/clone="132794"

BASE COUNT 97 a 56 c 95 g 111 t 5 others  
ORIGIN

Query Match 15.5%; Score 249; DB 5; Length 364;  
Best Local Similarity 96.6%; Pred. No. 0.00e+00;  
Matches 281; Conservative 0; Mismatches 5; Indels 5; Gaps 5;  
Db 1 GCATGGGAGAGACACAGAGTGAAGATGGAAGACAGGGAATTCAGTTGTAT 60  
QY 1315 gcatgggagagaccacagcagtgaggatgaagtggaagcagggaattcagttgtat 1374  
Db 61 CAAGGAACGTGATGCTACCAAGCATCATTTTGAAGCAGACACGTGGCAAGGCAAAACC 120  
QY 1375 caaggaactgatgctaccaaagcatcattttgaagcagaacgtggcaagggcaaaacc 1434  
Db 121 GCGAAATCGCAGTGATGGGCTGCTGCTTTTCAGGCTTATGTCAGATAGCCTTTTA 180  
QY 1435 ggcgaatcgcagtgagtgctgctgcttttcagccttattgcccagatagcctttta 1494  
Db 181 TCTGTGGGATGACTGAATGTACTATCTTATATTTGACTTTGTATGTCAGTTCCTGGT 240  
QY 1495 tctgtgg-annnctgaatgtactatttataattgactttgtatgcagttccctggt 1553  
Db 241 TTTTGTGATATTGCATCATAGGACCTCTGGCATTTTAGGAATTACTAG 291  
QY 1554 ttttttg-atattgcatcatagg-acctctgg-cattttaa-aattactag 1600

Search completed: Sat Apr 24 23:15:07 1999  
Job time : 2372 secs.